

# Towards a greenhouse convention

There are encouraging signs that governments are willing to negotiate a convention to abate the emission of greenhouse gases, but the pitfalls should not be underestimated.

ALTHOUGH the greenhouse effect (or the threat thereof) may be the best thing to have happened to atmospheric and climate research for a long time, politicians seem wisely ill-prepared to wait for the jury to declare itself on the scale and even the reality of the phenomenon. For the past several years, it has been clear that the research, however intricate, will be capable of yielding unambiguous results long before it will be possible to negotiate an international convention without which the accumulation of carbon dioxide and other greenhouse gases in the atmosphere could be halted. Perhaps the most serious danger now is that committed governments will go rushing off in different directions, impeding rather than assisting the process of negotiation. An international workshop has already been arranged for Washington in October where government officials will at least attempt to define the diplomatic problems that have to be solved. But other governments are pushing for the creation of a new international organization, while still others would bless the United Nations Environment Program (UNEP) with extra cash and, presumably, regulatory power. Yet no amount of tinkering with committee structures will tackle the underlying problems, which are squarely political.

Suppose, to simplify the argument, that the release of refrigerants into the atmosphere has been contained by the Montreal Convention, which restricts the production of chlorofluorocarbons (CFCs), leaving carbon dioxide and methane as the greenhouse gases of concern. On the pattern of the Montreal negotiations, there will be talk in Washington in October of a convention binding industrialized countries either to hold their emission of CO<sub>2</sub> constant from some point in the future or to reduce them by some proportion. The trick, again on the Montreal pattern, will be to win agreement on concerted action without defining too clearly what that should be.

But governments that have been willing to fall in with restrictions on the use of CFCs will be far less ready to accept similar rules for economically much more important CO<sub>2</sub>. To the extent that restriction will be costly in one way or another, on this occasion, the industrialized countries will be saying that the most prosperous among them should be the first to make sacrifices. Similarly, and with reason, the poor countries of the world — some developing and some not — will howl that restrictions on CO<sub>2</sub> emission are restrictions on economic development as such. Appropriately, the World Bank and its relations

are to be prominent at the October meeting, but the occasion may simply serve to show that their resources are entirely inadequate for sustaining the development of the poor countries without increased use of fossil fuel.

There are also horrendous problems of enforcement within national jurisdictions to be dealt with. CFCs, made by a few identifiable chemical manufacturers, will seem by comparison to have been a simple problem. In advanced societies, the only feasible way of restricting consumption of materials as widely used as coal and petroleum products without creating artificial gluts and shortages is by taxation. That is the obviously preferred course because the market would then shift economic systems in the direction of fuel efficiency, blunting the edge of the demand for alternative sources of energy such as nuclear power. But what will then happen to those who keep warm by burning wood which they collect themselves, or who carelessly set on fire the buildings in which they live? Still further pitfalls lie in the path of the Washington workshop, not the least of which is the political interaction between CO<sub>2</sub> and methane, the measured concentration of which in the atmosphere is increasing relatively rapidly, but whose sources are relatively obscure. There are several possibilities, ranging from the intensive cultivation of rice in developing countries or of domestic animals in prosperous countries to managed waste disposal and deforestation. But while the relative importance of these sources is poorly understood, there is potentially endless friction in the prospect that countries which agree to lumber themselves with onerous taxes for the sake of restricting CO<sub>2</sub> then discover that their sacrifices are being undone by the unregulated production of methane elsewhere.

The issue of the rain forests, the Amazon in particular, is already a bone of contention, unlikely to be made less so by next October's workshop. Brazil naturally regards others' complaints at its management — some say mismanagement — of the Amazon basin as an infringement of its sovereignty. No doubt because the subtlety of the issue is becoming understood elsewhere, there has also grown up the notion that taxes raised on CO<sub>2</sub> in the industrialized world should be used to compensate the countries of the Amazon basin (Brazil is not the only one) and other such regions for self-restraint on forests. But that is the wrong way to manage a problem already delicate enough. If there is a compelling case for sterilizing



the forests of the Amazon, that should be done separately from the taxation of rich consumers, where the objective should be to restrain demand for fossil fuel. And if the rain forests are where they are because those places are productive of trees, might it not be better for the greenhouse problem that they should be used as sources of construction timber and then replanted quickly?

Complications such as these argue powerfully for leaving the greenhouse problem where it properly belongs, in the hands of responsible governments and the politicians who run them. Technical agencies of the United Nations such as UNEP (despite its political pretensions) and the World Meteorological Organization would quickly be swamped by the unfamiliarity of the task if they were saddled with it. That is why the greenhouse effect is best guarded against by the means already used to set up a regulatory framework for CFCs (the effectiveness of which has still to be demonstrated) — an *ad hoc* international convention. But this is not to say that there is no technical function to be carried out within the framework of a convention. As budgets for atmospheric research increase, the need for objective synthesis and analysis of climatic data and of guarded prognosis becomes ever more urgent. Those working for an international convention could do worse than by saying to those willing to join that, to begin with, there will be no costs beyond those of helping to support a first-rate hard-headed monitoring agency. □

## Beastly experiments

Regulation of the use of laboratory animals is necessary, but should journals censor what is published?

WHAT should be the policy of journals such as this on the publication of accounts of experiments that, while entirely within the law where they are carried out, may elsewhere be illegal or, at least, publicly offensive? This, for practical purposes, is the question raised last week by a correspondent writing from Britain (Scotland, to be specific) and commenting on an experiment with animals carried out in France (*Nature* 339, 248; 1989). Our correspondent, representing an organization called the Committee for the Reform of Animal Experimentation, gave it as his opinion that the experiments would not have been allowed by the legislation on laboratory animals in force in Britain since 1986 and that “no other reputable scientific journal published in this country [Britain] will accept reports of work that clearly would not have been authorized under British law”. The complaint is fairly, even temperately, made. It deserves an answer.

The most obvious weakness of the complaint is that it conflates two disparate issues — animal experimentation, which in most places requires regulation, and publication, which everywhere should be free from it (but, sadly, is often not). To separate the issues, think of (or, otherwise, imagine) a state in which public demonstrations against the government are illegal. Would it be reprehensible that

newspapers in that country should, of their own volition, decline to publish accounts of public demonstrations held in foreign capitals against foreign governments? Of course not. (That the governments concerned might decide that such reports would subversively put ideas into their own people's heads and would elect to impose involuntary censorship on its press is a different matter.) Nor could it be held that a scientific journal published in a country where genetic engineering is forbidden (West Germany, for the time being without a legal framework of regulation, is almost in that case — see page 327) would be acting improperly by publishing in that field. Lawyers would no doubt merely remark that the commission of illegal acts is criminal, but that describing of them is not.

The case of animal experimentation is more complicated. The days have long since gone when the research community claimed that the practice of research should be entirely free from regulation. Even if researchers had never neglected or mishandled animals, never used a laboratory animal when a tissue culture might do instead and had always set about the planning of experiments in which animals are involved with deliberation, care and the determination to cause the least pain, regulation would be necessary. In many countries, public concern about the use and the potential for the misuse of animals in research can be met only by a framework of regulation that is at once consistent and transparent. The beneficiaries are not merely laboratory animals and those concerned on their account, but researchers, whose work would be grossly encumbered if matters were differently arranged. Among reasonable people, the issues to be decided are how regulations should be drafted and applied.

That there should be variations of practice from one country to another, in relation both to the press and to animal experiments, is inevitable and unavoidable; legal systems differ. But even where, as in Britain, Japan and the United States (the three countries in which this journal is equivalently published), the press is free from formal regulation, it is wise that it should conduct itself in a seemly fashion. This journal's rule of thumb for at least a quarter of a century has been that the results of animal experiments that would not easily win general regulatory consent had better be of exceptional interest if they are to be published. This position is admittedly pragmatic (and requires a subjective judgement of what may be exceptionally interesting). The obvious counter-argument, that a refusal to publish will ensure that contentious experiments are not undertaken in the first place, might have more force if there were not, for other good reasons, such a variety of journals. □

### Swift retribution?

The French laboratory from which the disputed experiments were reported in *Nature* is one of the two INSERM units broken into on 20 May (see page 326). The experiments bear on the understanding of visual handicap in the newborn.



# Cold fusion gathering is incentive to collaboration

- Some evidence for neutron production
- Claims of energy generation mostly ignored

## Washington

THE latest conference on cold fusion, a three-day meeting organized by Los Alamos National Laboratory in Santa Fe, New Mexico, last week, lacked the drama and intensity of other meetings around the United States earlier in the month. "Relaxed and civilized" was how one participant described it.

Debate centred on the reality and origin of the low level of neutron emission from an electrode of palladium in heavy water reported by Steven Jones and his colleagues from Brigham Young University. Not only were Stanley Pons and Martin Fleischmann absent from the meeting, but their claims of extraordinary heat output from cold fusion received little attention.

John Appleby, of Texas A&M University, repeated his group's assertion that they have measured excess heat in a number of electrochemical cells. Although the magnitude of heat output is comparable to that claimed by Pons and Fleischmann, it remains unclear whether this has anything to do with fusion.

Another Texas A&M group, led by Kevin Wolf, said they had found substantial amounts of tritium in some cells. But although the amounts they reported were 100 to 10,000 times above background, the accumulations were still considerably less than expected if the excess heat were due to deuterium fusions which generated tritium. Moreover, an analysis of a heat-producing palladium electrode by a research division of Rockwell International showed no accumulation of helium.

Because the measurements have been done by different groups, no single cell has been assessed for heat, tritium and helium, but the small quantities of fusion products suggests that at best only some fraction of the excess heat can be of nuclear origin.

Apart from the Texas A&M results, most of the discussion in Santa Fe concerned neutrons. The most intriguing new results were from Los Alamos National Laboratory itself. Howard Menlove announced that his group had detected low neutron fluxes, in bursts and spikes rather than the continuous output claimed by Jones, from a variety of experiments. In some cases, they had simply placed palladium and titanium, in the form of powder, sponge or metal shards, in high-pressure deuterium gas; in about 30 per cent of their experiments, sporadic neutron emission was observed.

Because they have used three neutron

detectors simultaneously, swapped dummies for real test samples and made background measurements, Menlove says his group's experiments convince him that under some conditions neutrons are indeed created by the interaction of certain metals with deuterium. But their success is only occasional, and they are now trying to understand how the form and preparation of their metal samples determines the outcome of the tests.

Menlove's results are broadly consistent with those from the Brigham Young group, but direct comparisons are hard to make. The Los Alamos neutron detectors are about 100 times more efficient than those used by Jones, so that the bursts that Menlove sees, which typically contain 100 or 200 neutrons, would have been registered by Jones as a single particle. Nevertheless, the average neutron production in the two experiments is similar.

Set against these positive signs were results from a comparable series of experiments carried out by a group from Yale University and Brookhaven National Laboratory. Announcing their findings, Moshe Gai of Yale had no neutrons at all to present, at a level of detection sensitivity considerably better than Jones's experiments. To resolve this contradictory state of affairs, Gai invited Jones to bring one of his neutron-generating cells to Yale. Jones accepted the offer.

The cooperative spirit evident at the meeting was counterpointed by the absence of Pons and Fleischmann, who said at a congressional hearing on cold fusion (*Nature* 339, 4; 1989), and have since repeated, that they were planning to work with scientists from Los Alamos in a critical assessment of their experiments. But a spokesman for Los Alamos made it plain that no such collaboration has occurred, because the University of Utah had not wished to enter into any agreement that it perceived would jeopardize its patenting and priority rights.

The lack of progress in substantiating the claims of Pons and Fleischmann has stalled any action in Congress. Staffers on the subcommittee on energy, research and development have been talking to several of the protagonists, including critics of cold fusion such as Nathan Lewis of the California Institute of Technology, and until a scientific consensus is reached, Congress, despite its enthusiasm of two weeks ago, will not be financing any large development projects.

David Lindley

## SCIENCE EDUCATION

### Hughes dollars draw students

#### Washington

THE Howard Hughes Medical Institute (HHMI) moved into the second phase of its programme of support for undergraduate science education last week with the award of \$61 million to 51 universities. The money will be used to support programmes designed to recruit and retain good science students.

HHMI is now by far the most biggest player in the field of US undergraduate science education, well ahead of the National Science Foundation, which has suffered a shortage of funds for its own initiatives. Legal constraints limit HHMI's role as a grant-giver: most of the income from its \$5,000 million is spent employing fully qualified researchers.

Two years ago, the institute awarded \$30.4 million to 34 liberal arts colleges and 10 historically black colleges (see *Nature* 329, 574; 1987). This year, the recipients are the best universities but the emphasis on encouraging women and minorities to study science remains. A total of \$22 million will be spent on intensive introductory and laboratory courses aimed at attracting students from these under-represented groups. The rest of the money is to be used for curriculum restructuring and new laboratory equipment (\$20 million), initiatives to involve more faculty scientists in undergraduate teaching and research projects (\$8 million) and outreach programmes aimed at students in local schools and colleges (\$11 million).

Penelope Austin

## GENETIC ENGINEERING

### The flowers that bloom next spring

#### Bonn

FORTY thousand genetically-engineered pink petunias are to bloom in West Germany despite the continuing controversy over new legislation to control such experiments (page 327). Last week, the Federal Health Office experiment approved an experiment by Heinz Saedler of the Max Planck Institute (MPI) for Breeding Research in Cologne (*Nature* 338, 194; 1989). But the MPI group will not plant the petunias until 1990; approval came too late for the 1989 season. Environmentalists have opposed the petunia planting, not because they thought the flowers harmful but because they thought it would set a precedent for releases of other genetically-engineered organisms.

Steven Dickman

#### Correction

IN A News article entitled "Fire strikes Jackson Laboratory" (*Nature* 339, 169; 18 May 1989), Gerald Callahan was incorrectly identified as being at the University of Colorado. He is at Colorado State University in Fort Collins. □



## AIDS test backlash

### London

THE Bulgarian government's compulsory AIDS screening programme has been attacked as undemocratic and a breach of human and civil rights by the Communist youth daily, *Narodna Mladezh*. The programme, introduced two years ago, imposes heavy fines on people who refuse to report for tests when ordered to do so and empowers the police to bring them in by force after a second refusal.

The first thrust of the programme was directed against 'high-risk' groups, in particular foreigners and Bulgarians who had spent more than a month abroad. As a result, the crews of Bulgarian ships returning home were marched off under armed guard to an "unknown destination" for testing. Journalists who questioned these tactics were asked to play such down such "mistakes" as being due to the tense situation that led to the programme, but the number of such incidents actually increased as the programme went on, according to *Narodna Mladezh*.

The screening process has now been extended to everybody in the capital, Sofia, between the ages of 16 and 65 — just at the time when the Bulgarian national assembly is introducing legislation intended to democratize life and increase the individual's human and civil rights. Many doctors, the paper says, are opposed to compulsory screening; Dr Svetoslava Popova, head of the epidemiology section of the directorate of preventative health care and state public-health control says: "I do not want to comment on the screening. I hope that the results will soon show that it is more appropriate to stop it."

The World Health Organization, *Narodna Mladezh* points out, does not recommend mass screening, on the grounds that it is inefficient as a means of halting the spread of the human immunodeficiency virus (HIV), and expensive. (Each test, it says, costs around US\$1.00 plus the cost of the disposable hypodermic and needle.) If the Bulgarian government has "serious arguments" in support of mass screening, why are they being kept secret? A proper explanation, the paper urges, would "surely serve as one of our weapons against the disease".

Unfortunately, *Narodna Mladezh* undermines its case by failing to clarify its data. It gives the number of people tested, as of 1 May, as 100,000 and says that there has not been a single seropositive case. If these figures are correct, they must refer only to the Sofia screening programme. Nationwide, more than 1.5 million people (citizens and resident foreigners) have been tested, and up to 15 March, 71 were HIV-positive.

Vera Rich

## Bombshell for life sciences

### Paris

A MAJOR overhaul of French life sciences research is under way which could result in several research groups being closed down and resources concentrated at "centres of excellence". Last year, when French research minister, Hubert Curien, announced an increase in funding for the country's major civil research organisation, the Centre National de la Recherche Scientifique (CNRS), he made it clear that some "housekeeping" would be called for (see *Nature* 334, 579; 1988). Now, the recently nominated director of life sciences at CNRS, Claude Paoletti, has decided to apply Mr Curien's directive "exactly and with great determination".

For 1989, 284 new research posts were created within CNRS and laboratory running costs have nominally increased by 5.1 per cent. But in life sciences, which absorbs about a quarter of the FF9,700 million (\$1,500 million) CNRS budget, this increase, for existing research groups, has been whittled to at most 0.6 per cent, once inflation has been taken into account. Some laboratories are even complaining that they have less money than last year. One group says that only 40 per cent of their research funds now comes from their 'employer', CNRS.

Although he criticizes the underfunding of CNRS life science research, Paoletti's 'new broom' approach is unequivocal and is dominated by two "absolute and unavoidable rules". First, researchers who are "lax" are likely to find their grants cut.

In the editorial of the latest CNRS life-sciences newsletter, Paoletti says that

## Animals released

### Paris

Two laboratories of the French national institutes of health and medical research (INSERM) at Lyons were broken into during the night of 20 May and more than 70 animals removed. About 40 macaque monkeys, 20 dogs, as well as cats, rabbits and marsupials were taken, apparently causing some harm to the animals, according to an INSERM communique. The thieves, whose motives are still not known, also stole notebooks from the two laboratories — the vascular surgery and organ transplant unit directed by Mari-Rose Eloy and Marc Jeannerod's department of experimental neuropsychology.

The theft, says the communique, is "catastrophic" for the researchers, who were investigating newborn motor and visual handicap, Alzheimer's disease and cardiovascular prostheses. The affected laboratories conformed to the latest strict European regulations for animal experiments.

Peter Coles

evaluation — in terms of publications, visits from foreign researchers, number of doctoral students and the number of patent applications — will ensure that "unjustified grants" will not be awarded. At a meeting of laboratory directors last week, Paoletti said that "some disciplines are dead. To inject more funds would be a waste of money". The CNRS only has a statutory duty "to provide researchers with a salary and a job, not the means to carry out research".

Second, he wants to "concentrate financial means on a diminishing number of research structures and on a restricted but better equipped and less dispersed population of researchers" — centres of excellence. The 'housekeeping' has already started. In January this year, Paoletti explained that the 306 laboratories would initially only receive 90 per cent of last year's budget and would have to apply for the rest. "Seventy per cent of laboratories", he says "will get their 10 per cent plus an extra 2 per cent. Some will get more than this, either because they are excellent or because they are working on urgent problems, or because they respond to a need for regionalisation". About 45 laboratories will end up with less than last year. "These are laboratories which I intend to demote and even, eventually, to close".

According to one laboratory director, researchers in life sciences are "very, very worried" about the new moves. The "lame duck" policy will hit hardest at the 227 so-called "associated laboratories" which operate within university departments rather than in autonomous centres. As many of these laboratories are in provincial areas, closures could leave regional universities with even less research capability in biology than at present — bad news for a government pushing a policy of decentralization. Curien does not favour the idea of centres of excellence, although he believes academically weak research departments should be closed.

Aware that cuts are unpopular, Paoletti has launched a campaign to claw back about FF1,000 million of the life-sciences budget automatically syphoned off each year to pay for expensive instruments, such as astronomical telescopes, that biologists do not use. Ten per cent of the global CNRS budget is reserved for these instruments, a 'tax' which is levied irrespective of discipline. Paoletti is adamant that this is unjust, especially as colleagues in the national medical research institutes do not have to pay a similar levy.

At the Paris meeting, Paoletti dropped several bombshells, including a suggestion that laboratory annual reports should be written in English. He made it clear that if his overall policy was not acceptable, he would resign.

Peter Coles

## GENETIC ENGINEERING

# Germany edges towards law

## Bonn

REPRESENTATIVES of some 25 interest groups delivered their responses to a draft 'basic law' to regulate genetic engineering at a closed hearing in the West German Health Ministry last week.

The government has made the ambitious promise that it will pass the completed law by the end of 1990, when a new parliament will be elected. But basic research scientists, worried about the potential consequences of the law, are pressing the government for changes before it takes effect.

Science organizations, trades unions, industry and ecology groups were represented at the ministry meeting. Their views will be taken into account when the ministry presents the draft to the cabinet, which is expected to approve it by 27 June and send it to the upper house of parliament, the Bundesrat, for a vote. The lower house, or Bundestag, will begin debating the law in the autumn.

The new law, which has been in discussion for nearly five years (see *Nature* 336, 611; 1988), is a "framework" for more detailed industrial production using recombinant DNA and the release of genetically modified organisms into the environment.

The law will give legal status to the current mechanisms for self-regulation of genetic engineering. These mechanisms include a central body (the "Central Commission for Biological Safety" or ZKBS in German) to evaluate applications to perform recombinant DNA experiments; four "security levels" for experiments with pathogenic bacteria; and regulations for release. Matters newly covered by the law include a provision for public participation in the licensing procedure and criminal penalties for breaking the rules.

Basic research organizations have begun reluctantly to accept that the law is inevitable and are trying now to shape the legislation to defend their interests.

Researchers object most strongly to two aspects of the draft law: the regulations concerning pathogenic bacteria and the composition and role of the ZKBS. The draft calls for the registration with the *Länder* (states) of "non-pathogenic organisms" and "safety strains" which do not present a risk to laboratory personnel or the population. Pathogenic organisms that present a moderate to significant risk must be individually registered with the Federal Health Office.

Ernst-Ludwig Winnacker, vice-president of the Deutsche Forschungsgemeinschaft (DFG), pleaded at the Health Ministry hearing for a removal of registration for the first group of organisms and for a general — rather than individual —

licensing procedure for the least dangerous of the second group. Some laboratories would need to file "fifteen hundred applications a year" to meet this requirement, said Winnacker, a situation that would be "ridiculous".

Researchers also objected to the stated purpose of the law: to protect people and the environment from "the dangers of genetic engineering". Klaus Fleischmann, Executive Secretary of the organization of German National Research Centres (*Grossforschungseinrichtungen*) protested that the dangers of the environmental release of genetically modified organisms are no worse than the dangers of the release of their unmodified precursors.

Both Winnacker and Fleischmann favour a centralized licensing procedure carried out by ZKBS rather than by the *Länder*. Most of the cabinet are likely to support them in this, but they will probably be opposed by the Federal Environment Ministry as well as the

## DNA FINGERPRINTS

# Profiles bank on the way

## London

A NATIONAL register of DNA profiles may be created within 2 to 5 years, according to the British government's Home Office. Responding to a report from the House of Commons Home Affairs Committee on the national Forensic Science Service, the government announced last week that work is now under way in the Central Research Support Establishment to prepare a computer program suitable for a database of DNA profiles. By the time this is ready, it is hoped that consensus may have been reached on a standard technique for DNA fingerprinting and on the legal and ethical questions relating to the creation of the database.

Since 1987, when DNA fingerprinting was developed by Professor Alex Jeffreys of the University of Leicester, the technique has been used in 200 court cases leading to criminal conviction. The technique is carried out for the criminal courts only by the Forensic Science Service. ICI, the only private laboratory to use the technique, can do so only for civil cases, such as immigration or paternity disputes, as well as for overseas police forces.

At present the law requires that an individual must give consent before a sample can be taken for use in a DNA test. A police officer of at least the rank of superintendent must authorize the taking of the sample, and only if there are grounds for believing that the accused person has been involved in a serious arrestable offence and that the sample could confirm or disprove his involvement.

*Länder*. Environment Minister Klaus Töpfer (Christian Democrat) has already objected to the draft law as it reduces the regulatory power of his ministry.

The *Länder* are expected to recommend changes in the law when it reaches the Bundesrat. In addition to trying to gain as much regulatory power as possible, the *Länder* are expected to insist, as will the opposition Social Democrats in the Bundestag, that there be more public participation in the licensing procedure.

West German industry, which has been pushing hard for the law, is basically satisfied with the draft, as is the trade union — IG Chemie — representing chemical industry workers. The union fears that many jobs will be lost if the biotechnology industry continues the exodus from West Germany that it fears has already begun. Other unions represented by the umbrella labour organization DGB (*Deutsche Gewerkschaftsbund*) take a much more sceptical view of the law, taking a position one observer at the Health Ministry hearing called "very Green".

Steven Dickman

But the committee argues that the law should be changed. "Provided the level of invasion of privacy is acceptably low", it said, there is no objection to making it a legal requirement to provide a sample for a DNA test.

The forensic service, which runs six laboratories on a budget of £12 million a year, was severely criticized in the Home Affairs Committee report, which found it remarkable that the scientific standards and reputation of the service remain high, since it "had fallen on hard times". It could not meet the demand for its work, was losing some of its best staff, and morale was at rock bottom.

The government accepted the criticisms, but says that "not all can be quickly solved". Since publication of the report in February, though, "a good deal has been done to meet the problems identified". A new director-general has been appointed, staff increases are planned and pay has been improved, says the government.

The government has made no commitment to increase resources for the overstretched Forensic Science Service. Instead, it says that the service should be funded not directly through the Home Office but through the police, who would be able to judge the value of the Service to them and thus the resources they are prepared to commit to it. The service is also under consideration for the status of an "executive agency", which would make it more independent from government, and free to carry out private work.

Christine McGourty



# US explores tritium options

## Washington

THE unintentional moratorium on tritium production in the United States is now expected to last until at least next year, when Department of Energy (DoE) officials hope one of the reactors at the Savannah River Plant in South Carolina can be restarted. The hiatus has bolstered arguments that Congress must support DoE plans for two new tritium production reactors to replace the aging Savannah River facilities. But an *ad hoc* group of scientists and arms control negotiators see the moratorium in a different light, and last week they told a Congressional committee that whether intentional or not, a moratorium could facilitate moves towards strategic arms control without seriously damaging the present US nuclear arsenal.

Tritium is used to boost the yield and lower the weight of nuclear weapons. Because tritium decays in 12.5 years, maintaining a nuclear stockpile inevitably means recharging weapons from time to time. But without the production reactors at Savannah River — all three of which are now shut down for safety reasons — the United States has no indigenous capacity for supplying new tritium for military purposes. Government officials say there is already a "crisis" for the US nuclear arsenal, and unless something is done at once the crisis will only deepen.

But George Rathjens, professor of political science at the Massachusetts Institute of Technology and Carson Mark, a consultant physicist at the Los Alamos National Laboratory where nuclear weapons are designed, say talk of a crisis is just so much smoke. Testifying last Tuesday before the House of Representatives Armed Services Committee panel on DoE defence nuclear facilities last week Rathjens pointed out weapons are initially supplied with more tritium than necessary to give them a "shelf life". Such an over-supply could be spread among the arsenal, postponing the time when a new supply would be essential. Additional surplus could come from tritium removed from weapons scheduled for decommissioning, or due to be dismantled by treaty.

Rathjens and Mark are among the authors of a tritium policy paper produced in collaboration with the Nuclear Control Institute in Washington. They argued that by making a commitment to new production reactors, the United States was sending the wrong signal on arms control. Why not, they argued, seize the initiative, using the *de facto* tritium moratorium as a lever to encourage serious movement on strategic arms talks. They urged Congress to continue preliminary design on new reactors, but to postpone construction decisions until serious arms control talks had been pursued.

Last year, DoE proposed to build a new heavy water reactor at the Savannah River Plant to supply 100 per cent of US tritium needs, and a second, backup facility in Idaho using a modular high-temperature gas reactor (MHTGR) design capable of supplying 50 per cent of estimated need (see *Nature* 334, 558; 1988). The two reactors were to be built over the next decade, at a cost of \$6,800 million. The Bush administration has requested \$300 million for the project in the 1990 budget.

But Mark scoffed at DoE estimates that a new reactor would take ten years to build, pointing out that some of the existing Savannah River plants were designed, built and began operating in less than three years. Mark argues that the Savannah River plants could be restarted and kept running long enough to prevent any critical depletion of the stockpile.

But Robert Barker, assistant to the Sec-

## HIGH-ENERGY PHYSICS

### Ambitious institute

#### Sydney

THE lack of suitable facilities necessary for advanced research into high-energy physics has prompted Australian physicists to form the Australian Institute of High Energy Physics (AUSHEP). This will link Australian scientists to the international consortium of leading particle-physics research centres.

The institute, a cooperative venture between the universities of Adelaide, Melbourne and Sydney, the Royal Melbourne Institute of Technology and the Australian Nuclear Scientific and Technology Organisation, aims to improve the level of graduate training and public awareness in high-energy physics as well as providing access to accelerators and associated detection systems overseas.

The universities of Sydney and Melbourne will be developing joint detector development laboratories and designing and testing modular microvertex detectors for the next generation of collider and fixed target experiments in particle physics. Australia already has a connection with the CERN UA2 detector project.

Detector development work will be contracted out to Australian industry. "The advantage to Australian industry is that it will be able to cooperate in the development and construction of high-level technology," says Professor Bruce McKellar, from the physics department at Melbourne. This, it is hoped, will also generate additional income for the firms involved.

The institute is hoping to receive additional financial support from the Federal Government's Australian Research Council up to A\$500,000 a year. Tania Ewing

retary of Defense for atomic energy, insisted there was a crisis, and that failure to begin work on new production reactors was tantamount to unilateral disarmament. Barker said that Pentagon officials were well aware that careful husbanding could stretch the life of existing weapons, although he declined to provide specific numbers in an open hearing.

At a second hearing the following day, DoE officials received criticism for their tritium production plans from another direction. According to J. Dexter Peach, assistant comptroller of the General Accounting Office (GAO), "the need for a new production reactor is more acute than it was [last August] when DoE made its recommendation to Congress". Peach says DoE seriously underestimated the delays it would face in restarting existing Savannah River facilities. GAO's analysis also suggests that it will be closer to 12.5 years before a new heavy water reactor can begin producing tritium, and at least as long for the MHTGR. Joseph Palca

## ANTARCTIC TREATY

### Australia says no

#### Sydney

THE Australian government will not sign a convention supplementing the Antarctic Treaty that would lay down rules for the exploitation of Antarctic mineral resources. Instead, Australia will push for the continent to be protected as a wilderness park. The convention will now lapse as its ratification requires all seven Antarctic claimant nations to sign the treaty.

The government's action is being seen by many as a move to woo the environmentalist vote which is becoming an increasingly powerful force in state elections. Until recently the Minister for Foreign Affairs, Senator Gareth Evans, and the Minister for the Environment, Senator Graham Richardson, had favoured the convention, claiming a mining treaty would be the most practical way to avoid unregulated and potentially disastrous attempts to exploit Antarctica. The Cabinet decision was also influenced by comments from the French Prime Minister, M. Rocard, that his Government would veto the convention for environmental reasons.

Australia will now try to convince other nations of the value of a wilderness park covering the whole of Antarctica and its surrounding oceans. This would be negotiated within the existing Antarctic Treaty which stipulates that Antarctica be used for scientific and peaceful purposes only.

The Prime Minister, Bob Hawke, will be meeting with Rocard in June and is expected to raise the Antarctic issue there and again at the next Antarctic Consultative Meeting in Paris in October.

Tania Ewing



## MILITARY ROCKETRY

# India's ballistic success

## New Delhi

INDIA last week signalled its entry into the small group of nations capable of building ballistic missiles with the successful test firing of the 'Agni' intermediate-range ballistic missile. Twice postponed because of technical problems, the launch took place on 22 May from the defence range in Chandipur on Orissa's east coast amid great publicity.

Except for a cryptic announcement that the test was a success, the government released few details of the flight. The Defence Ministry said the missile "followed the pre-determined path, impacted on the designated area in the Bay of Bengal and fully met the objectives of the mission". The Defence Research Development Organisation (DRDO), which developed the missile, claimed that the test was essential to test advanced technologies but said nothing about the primary aim of the mission: testing the first home-made re-entry heat shield, a key component of a weapons delivery system.

Agni is the third missile to come out of the \$500-million integrated guided-missile programme of DRDO. It is claimed to be almost fully indigenous, one reason why India could go ahead with its test in the face of opposition from the United States. The two-stage, 14-tonne 19-metre tall missile is said to be able to deliver a one-tonne warhead to a target up to 2,500 km away within 15 minutes. This range

includes some Chinese cities.

In 1987, India tested a surface-to-air missile, Trishul, and last year demonstrated the surface-to-surface missile, Prithvi, with a range of 250 km, sufficient to reach targets in Pakistan. It is no surprise that the launch of Agni evoked angry reactions from Pakistan, which called it a threat to regional peace.

The Indian Prime Minister, Rajiv Gandhi, has declared that India "is a non-violent country with no aggressive designs on anyone". In a statement issued immediately after the launch, Gandhi assured the world that Agni "is not a nuclear weapons system". He said "it provided an option to deliver non-nuclear weapons with high precision at long ranges" and that its test was "part of continuing efforts to safeguard our independence and security by self-reliant means". Dismissing allegations that Agni will be tipped with nuclear warheads, Gandhi recalled that India, which detonated a nuclear device 15 years ago, set an example to the whole world by refusing to convert this nuclear capability into nuclear weapons. "We wish to keep it this way", he said.

But some hawks would like India to go nuclear quickly in order to deter potential aggressors. As one defence analyst put it, Agni may be a major contribution to India's negotiating strength for advancing international peace and stability.

**K.S. Jayaraman**

## SPACE ROCKETRY

# Brazil's two-thirds VLS success

## São Paulo

ALTHOUGH still many steps behind India (see above), Brazil too showed that it is on its way to membership in the rocketry club with the successful launch of a reduced-scale version of its Satellite Vehicle Launcher (VLS) on 18 May. And, like India, Brazil is feeling the pinch as advanced nations try to prevent Third World nations from developing ballistic missiles by limiting the export of sensitive high-technology materials.

Brazil's four-stage rocket was launched from the Air Force base Barreira do Inferno ('Barrier of Hell') near Natal, in the northeastern state of Rio Grande do Norte. The rocket was a one-third scaled-down version of the planned 18-metre tall VLS rocket. The goal is to place a 120 kilogramme Data Collecting Satellite (SCD-1) into a 750 kilometre orbit. But despite the success of the test launch, the programme has been set back three years as boycotts and tight budgets take effect.

In April 1987, the group of seven industrialized nations banned the export of all materials critical to the development of rockets capable of delivering a 500 kilo-

gram payload to a distance of 300 kilometres. The supply of electronic components for rocket guidance was hit hardest.

The advanced nations' concern over the Brazilian programme is heightened by the fact that it is under the control of the Brazilian Air Force rather than a civil organization. The advanced nations suspect that the next step will be to go from a space programme to the development of a ballistic missile.

A complicating factor is that Brazil, despite its rabid anti-communism, has made deals with both China and the USSR to try to gain access to advanced technology. INPE, the civilian Institute for Space Research, already has an agreement with China to build a remote-sensing satellite.

The high-technology boycott has meant considerable frustration for INPE. Its SCD-1 satellite is due to be completed later this year. But the Air Force will not now complete its rocket until 1992 rather than the planned 1989. Despite complaints from some INPE members, the military insists on waiting for the indigenous VLS to be complete rather than hire a place on a foreign launcher. Ricardo Bonalume Neto

## TECHNOLOGY TRADE

# Japan warms to the Cray

## Tokyo

WHILE the US government considers new ways to put pressure on the Japanese government to make them buy more US supercomputers for its research institutes and universities, Cray Research Inc. says that Japanese industry has already taken to US machines with unbridled enthusiasm.

According to Cray, Japanese companies now own nearly as many supercomputers as US companies. Sumitomo Chemical Co. has now joined the club. But the sales may not please all sectors of US industry: the computers are used to help design world-beating high-technology products.

Japanese supercomputer sales are concentrated in the private sector. In the United States, nearly two-thirds of the 142 supercomputers currently in use are in national laboratories and universities, according to Takehiko Kato of Cray Research Japan Ltd. In Japan, the reverse is true — about 80 out of 120 machines are in private companies. And it is the private sector that is growing rapidly.

Many of the machines are being used 'in house' by Japan's supercomputer manufacturers (NEC, Fujitsu and Hitachi) but they are also widely used in other industries. All of Japan's 'big five' car manufacturers (Toyota, Nissan, Mazda, Honda, and Mitsubishi Motors) use Cray computers to design new car models. Nissan has five Cray central processing units, more than any other automobile manufacturer in the world, Kato says.

Toshiba Corporation used a Cray in the development of its 1 megabit chips, which are now dominating the world market, and Sumitomo Chemical now plans to use one of the US machines for the molecular design of new chemicals. Cash-rich Japanese service companies are also buying supercomputers. Yamaichi and Daiwa securities companies are said to have ordered machines. And the information conglomerate Recruit, which is currently embroiled in a huge political scandal, leases access to two Crays that it bought second-hand from the domestic telecommunications giant Nippon Telegraph and Telephone Corporation (NTT), another Cray customer.

**David Swinbanks**

■ Japan, together with Brazil and India, was last week added to the US Administration's "enemies list" as a country deemed to have engaged in unfair trading practices harmful to US exports. Specifics cited in the Japanese case were wood products, supercomputers and satellites. The list was issued under the controversial 'Super 301' provision of the 1988 Omnibus Trade Act, and requires negotiation on the 'offending' practices to be concluded within 3 years. □

## Neutron scattering

SIR—The statements that the EMBL outstation in Grenoble "[has an] uncertain future" and that "neutrons have not turned out to be particularly useful for biologists" in Peter Newmark's article on the European Molecular Biology Laboratory (*Nature* 338, 724; 1989) require some comment.

Neutron scattering is a specialized technique for which the appropriate instrumentation exists in very few places, the foremost centre being in Grenoble at the Institut Laue Langevin (ILL) which is supported by an increasing number of European states. Distinguished biologists from all over the world apply for beam-time at the ILL. Because of the number of applications, experiments have to be chosen very carefully, and to the usual set of questions asked about any proposal before it is supported is added the question: "are neutrons really the only way to obtain this information?" The "usefulness" of neutron scattering, therefore, would not appear to be very high if it were judged by criteria based on number of papers or number of biologists involved. On the other hand, there is no doubt that the state of knowledge in a number of important branches of molecular biology would be much less advanced today had it not been for neutron-scattering experiments (and similar results could not have been obtained in any other way!). Recent examples that immediately come to mind include the quaternary structures of the ribosome<sup>1</sup> and of the Tet-repressor DNA complex<sup>2</sup>, and protein dynamics<sup>3</sup>.

The locations of all the proteins of the 30S subunit of the *Escherichia coli* ribosome are now known following ten years of hard work (at Yale and Brookhaven) and this provides a sound basis for the interpretation of many other results. Similar work is under way on the 50S subunit (Berlin and Grenoble). In the field of protein dynamics, inelastic neutron scattering provides unique experimental information with which to test the widely used theoretical method of molecular-dynamics simulation. There are many more examples: membrane structures, protein hydration and stabilization, hydrogen bonding, protein and nucleic acid conformations within small complexes, viruses and chromatin<sup>4</sup>.

Because of the juxtaposition of national (CNRS, CEA, INSERM, Université Joseph Fourier) and international (ILL, EMBL, MPI, ESRF) laboratories, many of the complementary structural techniques of molecular biology are undergoing strong development in Grenoble, with the most sophisticated instrumentation available. In this context, we believe

that the future of the EMBL outstation is anything but "uncertain".

S. CUSACK  
B. JACROT  
R. LEBERMAN

European Molecular Biology Laboratory,  
6900 Heidelberg 1, FRG

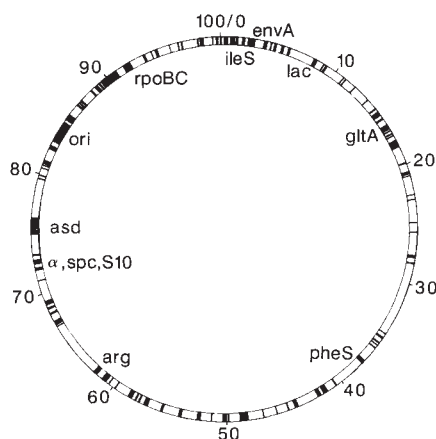
R. MAY  
P. TIMMINS  
G. ZACCAI

Institut Laue Langevin,  
156x, 38042 Grenoble, France

1. *Science* 238, 1403–1406 (1987).
2. *EMBO J.* 8, 1257–1263 (1989).
3. *Nature* 337, 754–756 (1989).
4. *Eur. Biophys. J.* 15, 257–268 (1988).

## *E. coli* genome

SIR—I read with interest Alun Anderson's report from Tokyo (*Nature* 338, 283; 1989) about the project to sequence in Japan the entire *Escherichia coli* genome.



Unfortunately, this report contains a misleading number, namely that about 450 kilobase pairs have been sequenced by researchers around the world. We have compiled these sequences as carefully as possible from both the GenBank and EMBL databanks and over a period of several years independently from the literature. Deleting the overlaps, we arrive at a total of 956,617 base pairs within 1,132 entries. This corresponds to about 20 per cent of the entire genome already sequenced by individual researchers. If you add a considerable number from unpublished material and some 2 per cent contributed by various insertion sequences, this number will be significantly higher. Our data are published in the sequence supplement of *Nucleic Acids Research* (25 April 1989). The included figure gives a visual impression of where the sequenced areas are located on the circular *E. coli* genome.

MANFRED KRÖGER

Institut für Mikrobiologie und  
Molekularbiologie,  
Justus-Liebig-Universität Giessen,  
Frankfurter Str. 107,  
D-6300 Giessen, FRG

## UK cell biology

SIR—The Southwood report will, according to your summary (*Nature* 338, 363; 1989) recommend that all biology in UK universities be carried out under the auspices of a single department containing two distinct sub-groups, the one molecular and the other traditional. I note with great concern that the subjects mentioned as falling naturally into one or other category fail to, and indeed cannot, include any aspect of cell biology.

A policy which does not naturally cover this field is short-sighted and misdirected because the cell is the environment of gene expression and the unit of biological function; its role is thus at the very heart of contemporary rather than traditional biology. The study of AIDS, cancer and mitosis, genetic disorders, immunology and developmental biology now involve taking an integrated view of the molecular genotype and phenotype in a cellular context and the work is done as much in biological as in medical departments. These examples also demonstrate that the policy will, if implemented as you report, lead to a fragmentation and hence to a diminution of teaching and research in fields that are becoming more amenable to investigation and hence more important as molecular biology advances.

The Southwood report apparently views biology as being either molecular or non-molecular, but this framework is already out of date. One effect of the molecular revolution has been that almost every branch of biology, be it molecular, cellular or traditional, has a strong molecular component. If therefore we are to plan for the future on the basis of a two-pronged approach to biology, it would probably be sensible to have one focusing on cell and molecular biology and the other on traditional subjects, with the expectation that many groups in both areas will actually be using molecular techniques.

JONATHAN BARD

26 Howard Place,  
Edinburgh EH3 5JY, UK

## Heliocentric Oresme

SIR—It is generally forgotten that the man who anticipated Copernicus<sup>1</sup> was Nicole Oresme who was born at Caen in France circa 1320.

In his *Questiones de spera*, c. 1340–50, he treats of a translatory motion of the Earth. Students of the problem may care to consult Balch's paper *The Laws of Oresme, Copernicus and Gresham*<sup>2</sup>.

FRANK W. COUSINS

43 Emanuel House,  
18 Rochester Row,  
London SW1, UK

1. *Nature* 338, 456 (1989).
2. Balch, T. W. *Proc. Am. Phil. Soc.* 47, 21 (1908) (reprinted).



# Prospects for a vaccine against HIV

Gordon Ada

Early optimism that a vaccine to prevent HIV infection would shortly be forthcoming had waned by the time of the fourth annual international conference on AIDS last year. On the eve of the fifth conference, are prospects any brighter?

VACCINATION is a form of immunoprophylaxis in which all or part of an infectious agent is administered once or a few times to generate an immune response that protects against subsequent infection. Many vaccines to prevent or control viral infections are highly successful. The widespread application of those composed of live (infectious), attenuated (non-pathogenic but immunogenic) virus, such as measles, mumps and rubella and polio viruses have reduced these diseases to minor public-health problems, at least in developed countries. Several attenuated viral vaccines under development, such as rotavirus, also show great promise. Some inactivated virus (rabies, polio) vaccines and a subunit (hepatitis B surface antigen) vaccine are fairly successful.

In view of this impressive record, it seems surprising that all attempts to make a practical vaccine against the human immunodeficiency virus, HIV, have so far been unsuccessful. A successful vaccine is the most direct way of showing that it is possible to stimulate an immune response in humans that will prevent or clear an HIV infection and thus prevent AIDS. The history of epidemics shows that although the mortality rate in an infected population may be very high (for AIDS patients in the United States, it is now more than 50 per cent), some people should unaided overcome the infection and survive. One finding is encouraging. In a cohort of 1,000 HIV-seropositive people in Baltimore that have been studied for some time, three reverted and have been seronegative for about 30 months<sup>1</sup>. This suggests that the active infection has been cleared in these people so that they are now immune. It is also possible that an attenuated form of HIV has evolved. Should HIV be experimentally attenuated for use as a vaccine?

Several properties of HIV complicate the development of effective and safe vaccines<sup>2</sup>. HIV is a retrovirus, which means that its RNA genome is copied to form complementary DNA, some of which integrates with the cellular DNA. Even though the cell may not initially produce infectious virus, it is thought to remain infected indefinitely (a latent infection), and possibly may be activated later to produce virus that could be subject to mutation. Naturally occurring superinfection with another retrovirus might

allow recombination to occur and a more virulent strain to be produced. Most people agree that this could happen if attenuated virus was used as a vaccine, so this otherwise attractive option has been shelved. This property of HIV also stresses the need for a vaccine which would prevent infection and hence stop the development of a latent infection.

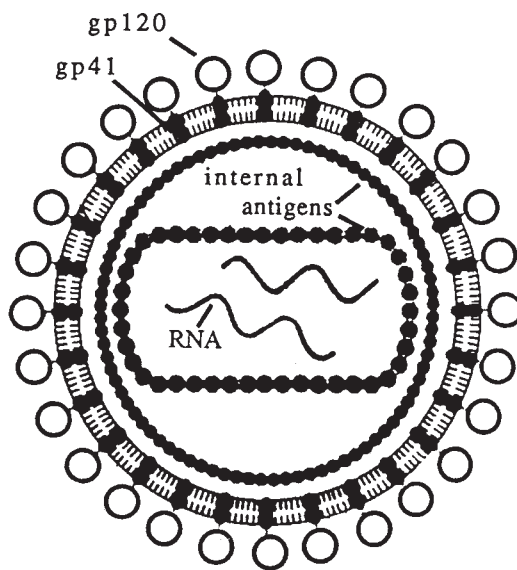
Another approach, the use of inactivated HIV as a vaccine, is being tried but is not widely favoured because there is a poor record of such vaccines against retroviruses. This type of immunization often fails to induce the production of cytotoxic T lymphocytes, cells which kill virus-infected cells as part of the normal

One particularly disturbing finding is that there is a great amount of antigenic variation in gp160 — at least 25 per cent of the amino acids may change<sup>3,4</sup>. A range of antigenically distinct HIVs has been isolated from one infected individual<sup>5,6</sup>. In the case of other viruses, marked antigenic drift in the protective antigen(s) of a virus has been a barrier to the development of effective vaccines (influenza, rhinoviruses). In HIV, one such segment occurs as a disulphide-bridged loop containing 18 amino acids, most of which are variable. In animal experiments, it has been found that this loop is the immunodominant site recognized by antibodies which neutralize the virus and thus prevent infection. One hopeful sign is that there seems to be a pattern of variation<sup>5,6</sup>; if so, a few sequences might give a broadly protective response. In sharp contrast, there is some experimental evidence that a conserved (non-variable) amino-acid sequence, present on a part of gp160 in the virus particle beyond but near the inner surface of the viral lipid membrane, is recognized by neutralizing antibody from infected people<sup>7,8</sup>, a finding that may be unique. It is encouraging that the less sensitive assays used previously for assaying viral infectivity have been replaced by quantitative techniques, allowing surer detection of important target sequences (called epitopes)<sup>9</sup>.

An antibody against gp120 may also have other functions. One useful role is to facilitate cell-mediated lysis of already infected cells (antibody-dependent cellular cytotoxicity). It has been reported that seropositive

patients who have not progressed to disease may have high titres of this antibody. Antibody may also accelerate the uptake of virus into cells such as macrophages and monocytes that have receptors recognizing antibody/virus/complement complexes and where some viruses are destroyed. However, HIVs productively infect these cells and this may be the way the virus is transported to different parts of the body, including immunoprivileged sites such as brain, epididymis and possibly bone marrow, which then become sanctuaries where infectious virus is continually produced.

Progress has been made in identifying epitopes recognized by both helper and cytotoxic T lymphocytes. At least one



Some features of HIV. The Env protein, gp160, has two components, gp120 and gp41, the latter projecting on both sides of the lipid membrane.

immune response. Furthermore, growing large amounts of HIV may be impractical.

Perhaps encouraged by the success of the hepatitis B vaccine, most effort is now directed to the development of a safe and effective subunit vaccine against HIV, based on the surface envelope antigen gp160, which is composed of gp120 and gp41 (see figure). To judge from many other viral infections, the vaccine should induce antibodies that neutralize viral infectivity as well as cytotoxic T lymphocytes. Both responses may develop, although often late during the natural infection. Much effort has been made to analyse in detail the properties of gp160, with both encouraging and disturbing results.



epitope is in a variable region of gp160. Findings from many other systems, such as malaria, indicate that a vaccine may need to contain several different T-cell epitopes to generate adequate immune responses in most recipients in an outbred population. Therefore, T-lymphocyte epitopes present in other viral antigens will need to be identified and incorporated into a vaccine. Several such epitopes have recently been described.

Two other properties of gp160 are of interest. The first is that there are similarities between several amino-acid sequences in the Env antigen and in some normal

Retention/deletion of biological active segments of HIV antigens in the construction of a 'synthetic' vaccine

#### Retain

Epitopes recognized by: neutralizing antibodies (gp160); antibodies involved in lysis of infected cells — ADCC (gp160); helper and cytotoxic T lymphocytes (gp160 and other viral antigens)

Peptide segments involved in binding to cellular receptors and in fusion to the cell membrane (gp160).

#### Delete

Epitopes recognized by antibodies which enhance infection of cells (macrophages, monocytes) (gp160)

Sequences showing homology with normal self components (gp160)

Sequences inducing suppression of immune responses (gp160)\*

\*Alternatively, it might be advantageous to include this sequence in the vaccine and generate a strong antibody response to it<sup>7</sup>.

host proteins. One of the latter is a class II HLA molecule, present at the surface of many host cells. Such findings raise concern that immunization with intact gp160 induces autoimmunity, possibly leading to autoimmune disease. The second finding is that a synthetic peptide representing a sequence in gp160 (in the transmembrane component gp41 — see figure) inhibits lymphocyte proliferation<sup>10</sup>. In HIV-seropositive individuals, the presence of antibody against this immunosuppressive peptide is reported to be associated with absence of disease<sup>11</sup>. This result may explain earlier reports that infected people have suppressor-cell activity which delays or prevents a protective antibody response.

A form of immunotherapy of considerable interest has the aim of inhibiting the interaction between the virus and CD4, the cellular receptor of HIV. One approach is to block the receptor itself, but as this molecule is involved in T-lymphocyte activation, this has the potential for undesirable consequences. Alternatively, 'soluble' CD4 molecules (or their equivalent) might be provided in sufficient amounts to bind to most if not all receptor-binding sites on the virus and so

prevent infection of susceptible cells. Daily injections of CD4 to monkeys infected with a simian immunodeficiency virus (SIV) related to HIV have recently been shown to have beneficial effects<sup>12</sup>. Methods have now been described<sup>13</sup> allowing such a reagent to be modified and remain in the circulation for a long time, thus requiring fewer administrations (which must be a requirement if this approach is to be feasible). As both CD4 and the receptor-binding site of gp160 are conserved, this approach bypasses the difficulties raised by antigenic variation. Specific antibodies (anti-idiotypes) mimicking the receptor molecule are also being investigated.

How can all this information best be used for the development of a vaccine? It is now possible to dissect a molecule to retain desirable and delete unwanted segments, or to modify selected features of it. The table indicates the activities that might be retained or deleted in the final antigen preparation. Furthermore, different presentation mechanisms could be used. Cytotoxic T lymphocyte responses might best be generated using a recombinant live vector (vaccinia, adeno, polio viruses), whereas administration with particular adjuvants, such as immunostimulatory complexes or liposomes, might give the best antibody responses<sup>14</sup>.

There is a great restriction on some HIV studies because of the lack of appropriate animal models. The recent description of an immunodeficient mouse strain which can be transplanted with human cells and infected with HIV offers the hope that susceptible animals will become available in sufficiently large numbers to test, at an early stage, different approaches to vaccine development.

There is one area that has received little attention. HIV is transmitted in four main ways — sexual intercourse, needle sticking, from mother to child and blood transfusion. All may involve free and/or cell-associated virus, and there is reason to believe that the latter may be an important mechanism. It is only recently that attention has been paid to defining the levels of free and cell-associated virus in semen. Not only is there great variation in cell numbers, in the degree of inflammation and in the amount of free or cell-associated virus, but semen also contains immunosuppressive factors<sup>15</sup>. The information required to develop the best vaccine will not be available until the state of virus in infected cells is characterized and the mechanism by which the virus infects the recipient is known.

Candidate vaccines are being developed and produced in the industrialized countries, where the responsibility of testing the product for safety and immunogenicity is generally accepted. For social and legal reasons there have been difficulties in the recruitment of

volunteers, but several phase I trials of putative vaccines are now under way. It seems probable that later trials (phase 2 and 3) will be carried out in developing countries. Furthermore, as vaccination is likely to be the only practical way to prevent and a major means for the control of HIV infection in the developing countries, it is critical that proper guiding principles are followed: beneficence (the obligation to do good rather than harm); justice (equal sharing among the population of the risks and of the subsequent benefits); and autonomy (subjects have freedom of choice about participating in research activities)<sup>16</sup>.

The great increase in knowledge about the virus, particularly the envelope protein, during the past year means that there are now more leads to follow and hence better prospects for effective, safe HIV vaccines. But there are still substantial obstacles. The Env antigen exists as a trimeric molecule and is enveloped in a 'shroud' of sugar molecules so that some peptide sequences are 'immunosilent'. An immune response to these may be useful. Some neutralizing antibodies recognize discontinuous epitopes<sup>9</sup>. If these are important, their structure needs to be determined. Many of the early successful vaccines were developed without much understanding about the properties of the virus or about how the immune responses required for protection were generated. Viruses such as HIV present a more formidable challenge.

Fortunately, there is always the prospect of an unexpected or novel result. The encouraging findings with a chimaeric poliovirus (see the paper by Almond and collaborators on page 385 of this issue<sup>17</sup>) containing a conserved epitope of gp160, which both reacts well with and generates human HIV-neutralizing antibody<sup>7,8</sup>, raises our hopes. Perhaps this is one vulnerable spot in HIV. □

Gordon Ada is in the Department of Immunology and Infectious Diseases, Hygiene and Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA.

1. Farzadegan, H. *et al.* *Ann. Int. Med.* **108**, 785–790 (1988).
2. Ada, G.L.J. *AIDS* **1**, 295–303 (1988).
3. Modrow, S. *et al.* *J. Virol.* **61**, 570–578 (1987).
4. Starcich, B.R. *et al.* *Cell* **45**, 637–648 (1986).
5. Saag, M.S. *et al.* *Nature* **334**, 440–444 (1988).
6. Fisher, A.G. *et al.* *Nature* **334**, 444–447 (1988).
7. Chan, T.C. *et al.* *EMBO J.* **5**, 3065–3071 (1986).
8. Dalgleish, A.G. *et al.* *Virology* **165**, 209–215 (1988).
9. WHO Workshop on the Measurement and Significance of Neutralizing Antibody to HIV and SIV. *WHO Bull.* (submitted).
10. Ruegg, C.L., Monell, C.R. & Strand, M. *J. Virol.* (in the press).
11. Klasse, P.J., Pipkorn, R. & Blomberg, J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5225–5229 (1988).
12. Watanabe, M. *et al.* *Nature* **337**, 267–270 (1989).
13. Capon, D.J. *et al.* *Nature* **337**, 525–531 (1989).
14. Ada, G.L. in *AIDS and the New Viruses* (eds Dalgleish, A.G. & Weiss, R.) (Academic, London, in the press).
15. Anderson, D.J. *et al.* in *The Heterosexual Transmission of AIDS* (ed. Alexander, N.) (Liss, New York, in the press).
16. World Health Organization. *Consultation: Criteria for International Testing of Candidate HIV Vaccines* (WHO, Geneva, 1989).
17. Evans, D.J. *et al.* *Nature* **339**, 385–388 (1989).

# AIDS and the new drug injector

S. R. Friedman, D. C. Des Jarlais, A. Neaigus, A. Abdul-Quader, J. L. Sotheran, M. Sufian, S. Tross and D. Goldsmith

Drug injection is one way in which human immunodeficiency virus is transmitted. Health programmes should take into account epidemiological differences between more experienced and new drug injectors.

INTRAVENOUS (IV) drug use has been the cause of one-third of the current AIDS cases in the United States and Europe, and the human immunodeficiency virus (HIV), the virus that leads to AIDS, has recently spread among IV drug users in developing countries such as Thailand, Brazil and Argentina<sup>1</sup>. New York City has the largest number of AIDS cases among IV drug users in the developed world: its AIDS caseload attributable to IV drug use<sup>2</sup> is roughly 7,000, comparable to the total number of cases in San Francisco, and half the total number of cases in Europe<sup>3</sup>. More than half of the estimated 200,000 drug injectors in New York City are infected with HIV<sup>3</sup>.

Efforts to stem the HIV infection rate among IV drug users around the world have concentrated on educating individuals about safer injection practices — such as not sharing injection equipment or cleaning shared syringes with bleach — and on the distribution of free clean needles. But our studies of IV drug users in New York City have led us to conclude that effective long-term prevention strategies must rest on a better understanding of the social interaction patterns of groups of IV drug users, and on accurate predictions of the changes the AIDS epidemic is likely to wreak on drug-use cultures.

Some have predicted that the AIDS epidemic will spell the end of illicit drug injection, as current and potential drug users switch to routes of administering drugs which do not involve the risk of HIV infection. Heroin and cocaine can be used intranasally, and both can be smoked (cocaine in its 'crack' or 'freebase' form, and heroin in an inhalation practice called 'chasing the dragon') to produce the same intense effects found with injection. But despite the fact that IV drug users in New York City have known since as early as 1984 that AIDS is transmitted through sharing injection equipment<sup>4,5</sup>, current studies have failed to show any measurable substitution in routes of drug administration. Increases in the number of persons smoking 'crack' cocaine have not been associated with decreases among those who inject cocaine<sup>6</sup>, and the number of new heroin injectors has not decreased since the start of the AIDS epidemic (B. Frank, personal communication). The knowledge of how AIDS is transmitted also has not deterred drug users from injection: a study of 83 heroin 'sniffers' in New York City found that more than 20

per cent went on to inject heroin eight months after participation in an educational programme<sup>7</sup>.

Research in London<sup>8</sup> and Amsterdam<sup>9</sup> also indicates that the AIDS epidemic has not prompted heroin injectors to switch to smoking. It may be argued that IV drug users in these cities are less aware of the risk of HIV infection because the number of AIDS cases among such drug users there is low. But other reports indicate that IV drug users in both cities are concerned about AIDS and know how to reduce their risks of infection<sup>10,11</sup>.

If people start injecting drugs despite the knowledge that AIDS can be transmitted through sharing injection equipment, it might be hoped that they would consistently follow safer injection procedures. Sadly, this does not seem to be the case. In a New York City study of drug injectors recruited from 'street' locations (in which drugs were being sold openly), Kleinman and colleagues<sup>12</sup> found that only very experienced IV drug users were taking deliberate steps to reduce their risk of AIDS. Only 16 per cent of users who had been injecting for two years or less had changed their injection behaviour to reduce their risk of HIV infection, compared to 29 per cent of those who had been injecting for 3–5 years, 33 per cent of those who had been injecting for 6–10 years, and 66 per cent of those who had been injecting for more than 10 years.

Surprisingly, our epidemiological data indicate that the low proportion of new IV drug users in New York City who have deliberately adopted safer injection practices does not translate into a higher prevalence of HIV infection. The table presents data on HIV infection from two of our continuing studies of street users in New York City. Both sets of data show a strong relationship between length of injection history and the likelihood of infection with HIV. Similar research in

Amsterdam<sup>11</sup> and Baltimore, Maryland (D. Vlahov, personal communication) has also shown that IV drug users have a greater chance of becoming infected with HIV the longer they inject drugs.

Why are so few new IV drug users infected with HIV, if only a handful are deliberately practising safer injection techniques? Two explanations are plausible: either new drug injectors share injection equipment less often than more experienced users, or they share equipment primarily with other new users, who are less likely to be infected with HIV.

There are indeed some differences in risk behaviour between newer and longer-term injectors. In both of the studies in Table 1, newer injectors were significantly less likely to rent previously used injection equipment. Renting used equipment — typically in 'shooting galleries' where users buy drugs and pay for the temporary use of syringes — is strongly associated with HIV exposure in New York City<sup>13,14</sup>, and provides a particularly effective mechanism for transmitting viruses across different social groups of IV drug users. But multiple logistic regression analysis indicates that these behavioural differences do not statistically explain the lower prevalence of HIV infection among newer IV drug users.

Thus attempts to explain the prevalence of HIV infection among new IV drug users by the behaviour patterns of individuals alone are probably inadequate. The groups with which IV drug users share equipment also affect their chances of becoming infected. If this interpretation is true, it will become increasingly important to study the social dynamics of groups of people who are starting IV drug use.

Typically, new IV drug users are initiated into injecting drugs by a friend or relative<sup>15</sup>. Because the first injection is usually unplanned, and new IV drug users are not psychologically committed to con-

Prevalence of HIV infection by years since subjects began to inject drugs

| Years injecting                | Five or less | More than five |
|--------------------------------|--------------|----------------|
| Bronx, Brooklyn, Queens        |              |                |
| Negative                       | 14           | 120            |
| Positive                       | 4            | 140            |
| Per cent positive              | 22           | 54             |
| $P(\text{chi-squared}) < 0.01$ |              |                |
| Lower East Side of Manhattan   |              |                |
| Negative                       | 17           | 88             |
| Positive                       | 4            | 94             |
| Per cent positive              | 19           | 52             |
| $P(\text{chi-squared}) < 0.01$ |              |                |





**Taylor & Francis**

London • New York • Philadelphia

## AIDS: Social Representation and Social Practices

Edited by **Peter Aggleton, Graham Hart and Peter Davies**, *Bristol Polytechnic, UK*

This book is the sequel to *Social Aspects of AIDS* published in 1988. It contains a selection of papers originally given at the Second UK conference on Social Aspects of AIDS and covers behavioural studies, health education and intervention, and representations of AIDS.

1989 1 85000 430 7 Cloth Price £18.95  
225pp 1 85000 431 5 Paper Price £8.95

## Social Aspects of AIDS

Edited by **Peter Aggleton and Hilary Homans**, *Bristol Polytechnic, UK*

This book brings together a selection of the papers originally given at the first UK conference on Social Aspects of AIDS. Central among the themes is the extent to which popular and professional understanding of AIDS have been informed by racism, homophobia and heterosexism.

1988 1 85000 363 7 Cloth Price £19.95  
200pp 1 85000 364 5 Paper Price £7.95

## AIDS: Principles, Practices and Politics (Reference Edition)

Edited by **Inge B Corless**, *University of North Carolina, USA* and **Mary Pittman-Lineman**, *Department of Public Health, San Francisco, USA*

AIDS, the acquired immune deficiency syndrome, is a complex and far reaching phenomenon that encompasses multidisciplinary principles, practices, and politics. This book addresses some of the most common, difficult and distressing issues confronting professionals and the general public. (An abridged, trade edition is also available see below.)

1989 0 89116 716 1 Cloth Price £58.00  
603pp 0 89116 979 2 Paper Price £30.00

## AIDS: Principles, Practices and Politics (Trade Edition)

Edited by **Inge B Corless**, *University of North Carolina, USA* and **Mary Pittman-Lineman**, *Department of Public Health, San Francisco, USA*

This comprehensive reference offers a thorough, up-to-date investigation of international impact of AIDS.

1988 0 89116 795 1 Cloth Price £25.00  
252pp 0 89116 772 2 Paper Price £10.00

(*Series in Death Education, Aging, and Health Care*) Reader Service No.12

For more information please contact:

Taylor & Francis, Rankine Road,  
Basingstoke, Hants RG24 0PR, UK  
Tel. (0256) 840366

tinued drug injection, new users almost never have their own equipment, and borrow equipment from the friend or relative. There are few data on the characteristics of people who initiate new IV drug users, but those which are available suggest that initiators are not experiencing problems related to their drug use, and are themselves likely to be relatively new to drug injection<sup>16</sup>.

According to this view, groups of uninfected friends and acquaintances start injecting over a period of some months, continue to inject primarily with one another, and only slowly do some of them begin to interact with the broader IV drug subculture. The Kleinman data<sup>12</sup> probably mirror this social process: lack of contact with more experienced injectors may well be the reason why newer injectors do not deliberately practise safer injection techniques. But the same lack of contact with more experienced injectors serves to protect new users from HIV infection. Ironically, contact with more experienced IV drug users increases their awareness of AIDS and safer injection techniques, but also increases their probability of sharing equipment with someone infected with HIV. In our study in the Bronx, Brooklyn and Queens recorded in the table, none of the 18 newer IV drug injectors — but 35 out of 260 older injectors — reported having shared injection equipment with someone who went on to develop AIDS ( $P < 0.15$  by Fisher's exact test).

Experienced IV drug users who remain uninfected now greatly outnumber newer injectors, so the need for efforts to encourage behavioural change among them is still urgent. The argument that AIDS prevention programmes will encourage or condone new IV drug use has been used to oppose the legal distribution of sterile injection equipment in the United States and Sweden, and the treatment of heroin addicts with the drug methadone in many European countries. However the available data show no increase in IV drug use associated with any AIDS prevention programme; indeed, participation in AIDS prevention programmes is often associated with reductions in IV drug use<sup>10,17,18</sup>.

But long-term reductions in the incidence of AIDS among IV drug users will depend upon ensuring that newer IV drug users know and practise safer injection techniques. This will be difficult. They may be extremely hard to locate, precisely because of their weak ties with the more accessible mainstream IV drug subculture. Fear-based health messages are likely to have limited effect, particularly with adolescents or young adults<sup>19</sup>. The difficulty of a fear-based approach will only be increased by the large amount of time that seems to elapse between users beginning to inject drugs and becoming infected with HIV, and the long interval

between infection and the onset of AIDS<sup>20</sup>. Programmes must be developed which make the need to avoid sharing possibly contaminated injection equipment real and legitimate to newer drug users.

Given the difficulties in developing programmes for reducing the risk of HIV infection among newer drug injectors, it will be important to learn why drug users start to inject drugs even though they know it could expose them to AIDS. More information is needed on the way new IV drug users think about risk and about taking drugs, and on how new injectors interact socially with each other and with more experienced drug users. The immediacy of these needs is compelling: there is some evidence that there may be an increase in the number of drug users who begin to inject drugs each year, if large numbers of 'crack' cocaine smokers turn to heroin use to moderate the 'down' that follows cocaine-induced euphoria. □

*Samuel R. Friedman, Alan Neaigus, Abu Abdul-Quader, Jo L. Sotharan, Meryl Sufian, and Douglas Goldsmith are at Narcotic and Drug Research Inc., 11 Beach Street, New York, New York 10013, USA. Don C. Des Jarlais is at the Beth Israel Medical Center, First Avenue at Sixteenth Street, New York, New York 10003, USA, and at Narcotic and Drug Research Inc. Susan Tross is at the Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA, and at Narcotic and Drug Research Inc.*

- Des Jarlais, D.C. & Friedman, S.R. *J. AIDS* **1**, 267–273 (1988).
- AIDS Surveillance Unit *AIDS Surveillance Update* (New York City Dept Health, April, 1988).
- Des Jarlais, D.C. et al. *J. Am. med. Assoc.* **261**, 1008–1012 (1989).
- Friedman, S.R. et al. *Int'l J. Addictions* **22**, 201–219 (1987).
- Selwyn, P.A., Feiner, C., Cox, C.P., Lipshutz, C. & Cohen, R.L. *AIDS* **1**, 247–254 (1987).
- Des Jarlais, D.C. & Friedman, S.R. *J. Am. med. Assoc.* **259**, 1945–1946 (1988).
- Des Jarlais, D.C., Casriel, C. & Friedman, S.R. Fifth International AIDS Conference, Montreal, Canada, 5–8 June, 1989.
- Gossop, M., Griffiths, P. & Strang, J. *Brit. J. Addictions* (in the press).
- Buning, E.C., van Brussel, G.H.A. & van Santen, G. in *Needle Sharing Among Intravenous Drug Abusers: National and International Drug Perspectives* (eds Battjes, R.J. & Pickens, R.W.) 59–74 (Government Printing Office, Washington, DC, 1988).
- Alldritt, L., Dolan, K., Donoghoe, M. & Stimson, G.V., Fourth International Conference on AIDS, Abstr. No. 8511, Stockholm, Sweden, 12–16 June, 1988.
- van den Hoek, J.A.R., Coutinho, R.A., van Haastrecht, H.J.A., van Zadelhoff, A.W. & Goudsmit, J. *AIDS* **2**, 56–60 (1988).
- Kleinman, P. et al. *Int'l J. Addictions* (in the press).
- Marmor, M. et al. *AIDS* **1**, 39–44 (1987).
- Schoenbaum, E.E. et al. Third International Conference on AIDS, Abstr. No. S34b, Washington, DC, USA, 1–5 June, 1987.
- Des Jarlais, D.C., Friedman, S.R. & Strug, D. in *The Social Dimensions of AIDS* (eds Feldman, D. & Johnson, T.) 111–125 (Praeger, New York, 1986).
- Zinberg, N.E. *Drug, Set and Setting* 127–132 (Yale University Press, New Haven, Connecticut, 1984).
- Jackson, J. & Rotkiewicz, L. Third International Conference on AIDS, Abstr. No. TH5.2, Washington, DC, USA, 1–5 June, 1987.
- Buning, E.C., Hartgers, C., Verster, A.D., van Santen, G.W. & Coutinho, R.A. Fourth International Conference on AIDS, Abstr. No. 8513, Stockholm, Sweden, 12–16 June, 1988.
- Botvin, G.J. in *Preventing Adolescent Drug Abuse: Intervention Strategies* (eds Glynn, T.J., Luekefeld, C.G. & Ludford, J.P.) 116 (Government Printing Office, Washington, DC, 1983).
- Bacchetti, P. & Moss, A.R. *Nature* **338**, 251–253 (1989).



# Who will see a nearby supernova?

The detection in 1987 of a supernova in the Large Magellanic Cloud has led to an estimate of the maximum rate of galactic supernovae so unrestrictive that the best response must be to run neutrino detectors for decades on end.

ALTHOUGH the supernova found in the Large Magellanic Cloud in February 1987 has kept many astronomers on their toes, it has not been much of a public spectacle. Even those of us lucky enough to be in the Southern Hemisphere have not been able to point to a new star shining in the sky. Indeed, even those with access to telescopes have had to use the most refined instrumentation to glean scraps of information from the supernova.

This is a far cry what it must have seemed like in China in 1054, when the star whose principal remnant is the Crab nebula blew up. Later, Tycho Brahe and Kepler each had the good luck to record supernova explosions in the Galaxy, which amounts to three visible galactic supernovae in a millennium. So can one work by the rule that there is one visible supernova every 300 years or so? The chances are probably greater than that, for the Crab explosion spotted by the Chinese appears not to have been recorded elsewhere in the world, even though it is said to have been bright enough to be visible by day. The rate of one explosion every 300 years, implying a (human) lifetime-chance of one in five, is probably a lower limit.

Now there is also what seems to be an upper limit, derived from the observation of the neutrinos accompanying the explosion of the Large Magellanic Cloud supernova (known as 1987a). Two years ago, it was reckoned to be something of a marvel that two special-purposes neutrino-detectors, one in the United States and one in Japan, had been functioning at the time and were able to record the small flux of neutrinos from the Large Magellanic Cloud at the time of the explosion.

The Irvine-Michigan-Brookhaven collaboration, which has for nearly a decade been operating a neutrino detector 600 metres underground in an Ohio salt-mine, has used just over two years' worth of data to fix an limit for the recurrence rate of galactic supernovae (S.T. Dye *et al. Phys. Rev. Lett.* **62**, 2069; 1989). The limit is that the galactic supernova rate must be less than 1.5 a year.

As these things go, the neutrino detector is a simple affair. Its purpose was originally to detect the decay of protons (still a live issue in high-energy physics although the lifetime of the proton is clearly greater than the least it might have been). There is a huge tank of water (8,000 tonnes of it) in which energetic neutrinos, by interaction with electrons or

protons, yield relativistic electrons or positrons which, by travelling faster than the velocity of light in water, yield pulses of Čerenkov radiation (light) that can be recorded by more than 2,000 sensitive photomultipliers.

Crudely, the greater the energy of the electrons into which the neutrino is converted, the more photomultiplier tubes will record signals, so it is possible to select events whose energy exceeds some threshold by requiring that the number of tubes which fire should exceed some number. (Some device of this kind is in any case needed to filter away random discharges of the photomultipliers.)

In reality, the operation of this deep-mine detector is plainly something of a well-regulated nightmare. Despite the great depth, the equipment records 2.7 cosmic-ray muons every second (which must be identified and discarded from the records by using the directional properties of the photomultipliers to show that these particles begin life outside the water tank). Cosmic-ray neutrino interactions within the water-tank are luckily less frequent, at roughly two a day.

Supernova 1987a blessed the neutrino-detector on 23 February 1987 with eight neutrino events in a time-span of 6 seconds. The similar Japanese experiment, Kamiokande-II, yielded 11 neutrino events at exactly the same time (within the one second uncertainty of the time measurement). Since those measurements, there can have been no doubt in reasonable people's minds that even very distant supernovae (the Large Magellanic Cloud is 50,000 parsec away) deliver neutrinos at the surface of the Earth.

So why not use the same records to search for signals from other supernovae? This is what the Irvine-Michigan-Brookhaven collaboration has now done. Briefly, they have found no other, but in the process they have been able to fix a limit on the presumed rate of supernova occurrence within the Galaxy.

The argument is both interesting and important. Galactic supernova should produce much stronger signals than 1987a. Roughly half the stars in the Galaxy are within 10,000 parsec of the Sun, which is also the distance of the galactic centre. So, in principle, searching the existing record for past galactic supernovae simply involves hunting for clutches of neutrino events in the Ohio records.

The practical difficulty, of course, is

hunting for bunches of putative neutrino events spanning a few seconds in records which, between them, span more than 700 days. The task has been slightly complicated by the need, at the Ohio detector, to deal with the two groups of records from before July 1984 and after May 1986 (when the sensitivity of the instrument had been improved).

By themselves, the separate records give modestly disappointing lower limits for the rate of galactic supernova explosions — there must be fewer than one every two years or so. Putting them both together gives the result that there are likely (with 90 per cent confidence) to be fewer than one every 1.5 years. The inference is relatively robust. The important provisos are that the likelihood of supernovae outside the 10,000 parsecs in which the first version of the Ohio equipment could detect explosions with reasonable efficiency is representative of the more distant parts of the Galaxy, and that the effect of the supernova is the collapse of a massive star either into a neutron star or a black hole.

Inevitably, the limit now obtained, based as it is on just two years of observation, is not much of a constraint on the theoreticians. Dye *et al.* point out that most estimates of the galactic rate of supernovae explosions, based on arguments from the pulsar population or the presumed rate of nucleosynthesis or even from direct observation of the supernova rate in other galaxies, give estimates of one every 10 years at the maximum. Wrily, the Irvine-Michigan-Brookhaven collaboration points out that it would take 70 years of observation with its detector to put a 90 per cent confidence limit on an estimate that the rate of supernova explosion is less than one every 30 years.

Quite what that means for the chance that some people now alive will see a supernova with the naked eye is therefore, for a long time to come, an open question. Some say that only three-quarters of the Galaxy is visible from the region of the Sun, but supernovae appear to be concentrated in regions of star-formation, where obscuring gas and dust will be common, so that the chances of seeing them will be even smaller. So Dye *et al.* are right to say that the best strategy for detecting supernovae will be to "build, maintain and operate for decades" detectors capable of surveying the whole Galaxy.

John Maddox

# Biological cold fusion

Gottfried Schatz

In eukaryotic cells, the proteins of most membrane-bounded organelles are initially inserted into a progenitor organelle (the rough endoplasmic reticulum) and only later distributed to their different destinations. In this sorting process, proteins destined for transfer are sequestered within membrane vesicles that bud off from a donor organelle and then fuse with the appropriate acceptor organelle (see figure). Three papers by Rothman and collaborators on pages 355, 397 and 398 of this issue<sup>1-3</sup> show that vesicle fusion in several distinct branches of this complex distribution network requires the same cytosolic protein. This suggests that most, and perhaps all, of these fusions are driven by the same machinery which becomes activated once a vesicle has docked onto its specific target membrane.

Biochemical analysis of these transfer steps became possible after Rothman and his colleagues<sup>4</sup> had worked out a cell-free system from Chinese hamster ovary (CHO) cells which catalyses the transfer of vesicles between different regions of the Golgi complex (see ref. 5 for a review). This transfer requires ATP, fatty acylCoA and at least two cytosolic proteins, one of which is inactivated by the sulphhydryl blocker *N*-ethylmaleimide (NEM). Subsequent work showed that the transfer involves several sequential steps which culminate in the fusion of vesicles with a medial Golgi sac; when the cell-free system is treated with NEM, vesicles bind to the target membrane but fail to fuse with it. Readdition of a cytosol fraction in the presence of ATP restores fusion. ATP and at least one NEM-sensitive cytosolic protein are thus necessary for the final fusion step.

Helped by this assay system, Rothman and his collaborators purified the NEM-sensitive protein (which they call NSF), characterized it as a tetramer with subunits of relative molecular mass 76,000, and raised a monoclonal antibody against it. This set the stage for the three studies reported in this issue.

The paper by Beckers *et al.*<sup>2</sup> reports that NSF is also required for transferring vesicles between the rough endoplasmic reticulum and the Golgi complex. Transfer requires ATP and is inhibited by NEM or the monoclonal antibody against NSF. NSF is required in a late, calcium-dependent transfer step; this step is most likely the fusion step. However, the published

data do not exclude the possibility that NSF is also needed at earlier steps.

Diaz *et al.*<sup>3</sup> find that this versatile protein also participates in the fusion of endosomes in a cell-free system derived from macrophages. Again, involvement of the protein can be deduced from the sensitivity of the process to NEM or to the anti-NSF monoclonal antibody, and from the finding that most of the inhibition by NEM is overcome by subsequent addition of NSF from CHO cells.

To get a closer look at this protein, Wilson *et al.*<sup>1</sup> cloned and sequenced its gene and find that the deduced protein product has 48 per cent sequence similarity with the product of the *SEC18* gene of

The sequence of mammalian NSF lacks any obvious hydrophobic regions resembling those of viral proteins which can induce membrane fusions by themselves. It is thus unclear whether NSF is a true 'fusogen' or whether it merely promotes the fusogenic action of some unknown partner protein. Whatever the answer, vesicle transport between membranes surely requires more than a single protein catalyst.

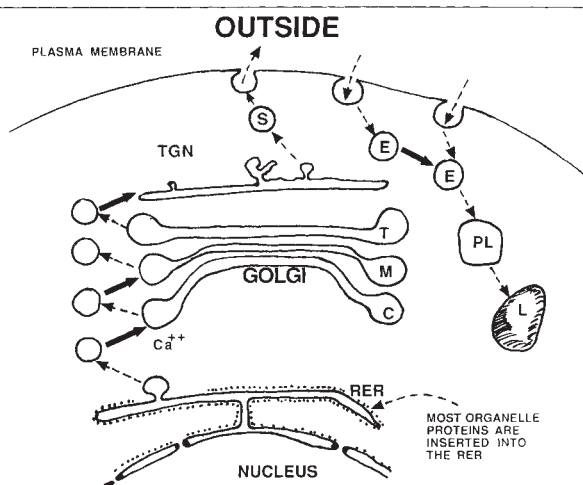
There is already evidence for a second cytosolic protein factor and for a membrane-associated NSF receptor, and the precise role of fatty acyl-CoA also remains to be clarified. Some of these questions will probably soon be answered as biochemical resolution of multicomponent systems is usually fairly rapid once the first component has been purified in an active state. This has now been achieved for the system discussed here.

The results from these new studies<sup>1-3</sup> also raise the possibility that fusions between different organelles derived from the rough endoplasmic reticulum may all be catalysed by the same set of proteins. These proteins might merely act sequentially or as a multisubunit 'fusion machine' that is assembled at its site of action in response to as-yet unknown signals. Either of these mechanisms would resemble the action of complement proteins whose sequential activation and assembly is triggered by specific immune reaction on the target membrane. If cells use the same instrumentation for most or all of these fusions, the task of unravelling the molecular orchestration of membrane flow will be made much easier.

The papers by Rothman and collaborators in this issue are remarkable in several respects. First, they exploit an impressive array of clever *in vitro* assays. Second, they underscore the fact

that a biological system can sometimes be simpler than originally surmised. Finally, they demonstrate the power of combining genetics with biochemistry, thereby providing yet another example of biological fusion. □

Gottfried Schatz is in the Department of Biochemistry, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.



In a eukaryotic cell, membrane vesicles that bud from a donor organelle and fuse with an acceptor organelle transfer membrane components and soluble vesicle contents between intracellular compartments, and between the inside and the outside of the cell. Membrane fusions now shown to be mediated by the NEM-sensitive protein (termed NSF in mammals and SEC18 protein in yeast) are indicated by thick arrows. RER, rough endoplasmic reticulum; C, M and T, *cis*-, medial- and *trans*-Golgi compartments; TGN, *trans*-Golgi network; E, early endosome; PL, prelysosome; S, secretory granule. (The scheme omits several pathways not relevant to this discussion.)

yeast. This gene had already been shown by Schekman *et al.*<sup>6</sup> to control the transfer of vesicles between the rough endoplasmic reticulum and the Golgi complex, and later studies suggested it has a function in endocytosis<sup>7</sup>. The deduced sequence of the SEC18 protein and its biochemical features are typical of a cytosolic protein<sup>8</sup>. Indeed, Wilson *et al.* show that a cytosol fraction from yeast can replace the endogenous NSF in the cell-free intra-Golgi transfer system from CHO cells; NSF activity is 10 times higher in a cytosol fraction from yeast cells that overproduce the SEC18 protein and undetectable in cytosol from cells bearing a mutant *sec18* allele. This last observation also suggests that, in yeast, SEC18 protein is the only functional NSF homologue.

1. Wilson, D. W. *et al.* *Nature* **339**, 355-359 (1989).
2. Beckers, C. J. M. *et al.* *Nature* **339**, 397-398 (1989).
3. Diaz, R. *et al.* *Nature* **339**, 398-400 (1989).
4. Glick, B. S. & Rothman, J. E. *Nature* **326**, 309-312 (1987).
5. Schatz, G. *Trends Biochem. Sci.* **10**, 95 (1985).
6. Schekman, R. & Novick, P. in *The Molecular Biology of the Yeast Saccharomyces cerevisiae* (eds Strathern, J. N., Jones, E. W. & Broach, J. R.) 361-398 (Cold Spring Harbor Laboratory, New York, 1982).
7. Riezman, H. *Cell* **40**, 1001-1009 (1985).
8. Eakle, K. A., Bernstein, M. & Emr, S. D. *Molec. cell. Biol.* **8**, 4098-4109 (1988).



## EARTHQUAKES

# Portents and predictions

Carl Kisslinger

How are earth scientists or public officials to evaluate predictions of damaging earthquakes and to decide whether measures to mitigate the effects are called for? Predictions of the place, time and magnitude of earthquakes have been made in several countries during the past two decades. Some have been verified<sup>1</sup>, some have been partially successful<sup>2</sup> and others have been false alarms. The basis of the prediction has been well-documented in some cases and not so clear in others. There is no adequate public record of false alarms or failures to predict from applications of methods for which successes have been claimed. Last year, two Soviet geophysicists, Keilis-Borok and Rotwain, and their US co-workers, Knopoff and Allen, presented in *Nature*<sup>3</sup> a prediction method based on pattern recognition that they have applied to the active zones of California. Keilis-Borok was also quizzed at the time by US experts on the approach. Although the panel concludes in its report<sup>4</sup>, just published, that there are no sufficient grounds for immediate action and that existing US methods should suffice, the Soviet method presents a useful new angle on the prediction problem.

Public officials need scientific guidance as to the level and types of responses that are appropriate when presented with a prediction of a strong earthquake in a populated area. In the United States, the National Earthquake Prediction Evaluation Council (NEPEC), established in 1979, advises the Director of the US Geological Survey who is responsible for the decision whether and when to issue a prediction.

Objective evaluation requires the adequate definition of what constitutes a prediction. It is generally agreed that a prediction should be a statement of the place, time and magnitude of a future earthquake, with some estimate of the uncertainty of each and sufficiently specific for the random occurrence of an event within the limits stated not to be highly probable, given the known seismicity of the region. Prediction, thus, implies evidence of a substantial increase in the probability of the specified event above the long-term value. Too few cases with all three parameters predicted have been completed for one to assign quantitative uncertainties either to the likelihood of occurrence or to each of the predicted parameters, even for the most promising methods.

Validation is especially difficult because the primitive scale of current methods means that every prediction is an experiment. Furthermore, prediction experiments often involve conditions that differ

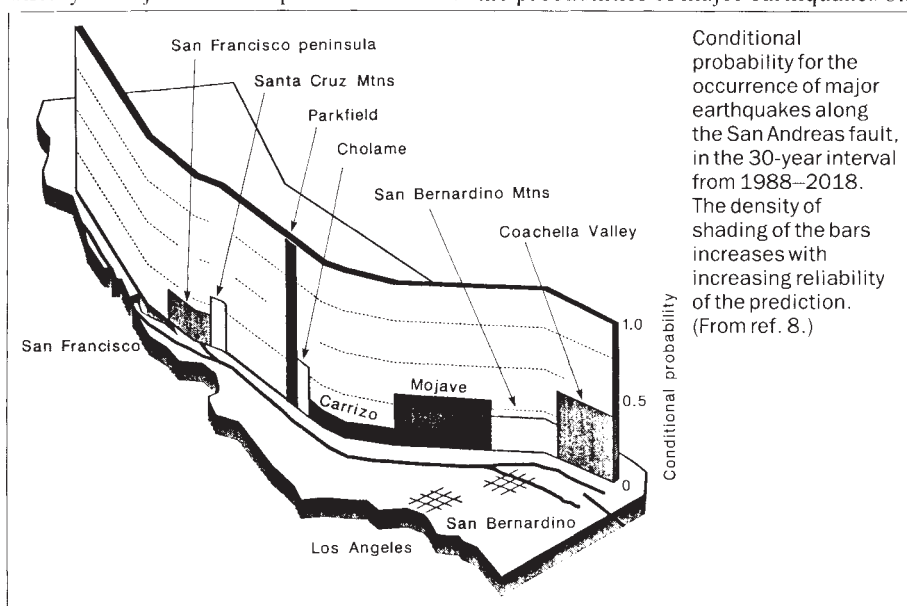
significantly from those in other tests: mode of faulting, position relative to tectonic plate boundaries, depth of the source and rate of tectonic strain accumulation all vary widely. In the absence of an adequate theory of seismogenesis, the effects of changes in any of these factors are hard to anticipate.

Predictions are conveniently classified, as long-term (a few years to a few decades), intermediate-term (a few weeks to a few years) and short-term (up to a few weeks)<sup>5</sup>. Different methods are used to formulate predictions in each category, and the socially useful purposes to which the predictions can be put also differ. Substantial progress has been made in the long-term problem, owing to improved understanding of the underlying tectonic processes and the reconstruction of the history of major events on particular faults

appearance of several might sufficiently increase the probability to justify a prediction.

The Parkfield Prediction Experiment<sup>7</sup>, on a section of the San Andreas fault in central California, is designed to determine if short-term precursors can be detected by a dense network of sensors, monitoring various geophysical quantities, and rapidly integrated to yield an accurate prediction. An intermediate-term prediction for the site has been formulated on the basis of the seismic history during the past 130 years and there is compelling evidence on geological grounds that this segment of the fault produces a "characteristic earthquake" at regular intervals of about 22 years. NEPEC judges the intermediate-term prediction for this site to be valid.

An example of contrasting approaches to long- and intermediate-term prediction is afforded by two recent examinations of the outlook for California. An *ad hoc* group, formed by NEPEC, has evaluated the probabilities of major earthquakes on



or in large fault zones for times longer than recorded history.

The most promising approaches to intermediate-term prediction<sup>6</sup> use analyses of the distribution of seismicity in a region in space and time and observations of localized crustal deformations. Wyss and Burford published in *Nature*<sup>1</sup> a successful prediction, based on seismic quiescence, of a moderate earthquake in central California, one year before it occurred. Short-term prediction turns out to be difficult. Initial hopes that a magic signal would be found have given way to the realization that an ideal precursor, some phenomenon that always precedes an earthquake and never occurs without one, apparently does not exist. It is more likely that precursory changes of several types occur, each of which indicates an increased probability of an event; the simultaneous

recognizable segments of the San Andreas fault during the next few decades<sup>8</sup> (see figure). The group calculated a conditional probability of occurrence of a strong earthquake on each fault segment considered, using the time since the last such event, the amount of the displacement in that event, and the best available information about the long-term slip rate. The resulting probabilities for the next 30 years, 0.5 for a magnitude-7 event in the San Francisco Bay area and 0.6 for one of magnitude 7.5–8 on the southern San Andreas fault, are described as minimum values of the hazard in the regions of California involved, because only the San Andreas fault was considered.

Another approach is to apply pattern recognition as described in *Nature*<sup>3</sup> last year by V.I. Keilis-Borok of the Institute of Physics of the Earth, Moscow, and

US Geological Survey



co-workers. One starts by examining a list of natural phenomena that might be expected to be associated with the occurrence of a strong earthquake; the list can be quite long. These phenomena must have features that can be given a quantitative, unambiguous definition. The occurrence or absence of each feature is noted during intervals preceding strong earthquakes and other intervals not followed by such events, for a catalogue of events in the region in question.

Each candidate is scored and those that emerge as diagnostic of dangerous or non-dangerous time intervals are retained as traits to be used for prediction. In recent work, traits have been derived from space-time patterns of seismicity. The algorithm is then applied to the current catalogue to evaluate the state of the selected traits and to search for times of increased probability (TIPs) of occurrence of an event above a specified threshold magnitude. The TIPs are arbitrarily assigned a duration of five years, but can be extended as warranted by the behaviour of the traits as time passes.

Such algorithms reveal TIPs for strong earthquakes in the California-Nevada region, but with poorly specified location. The independent finding of increased (but unspecified) probability of a large earthquake occurring in California before January 1989, alongside the results of American probabilistic estimates, led to a request from the Director of the US Geological Survey for NEPEC to evaluate the Soviet results. The published report<sup>1</sup> of the meeting in June last between Keilis-Borok and NEPEC shows that the panel had many difficulties in understanding the details of how the method works, the underlying assumptions and so on. Some were resolved. It is obvious that the purely phenomenological approach in which (by intent) no connection between a trait and the physics of the earthquake process is invoked, runs counter to the widely held view that a rational approach to prediction should be grounded in that physics. On the other hand, it was agreed that if traits that really work are identified, the research to learn why will inevitably follow.

During the NEPEC meeting, the status for California predictions was given by Keilis-Borok as: first, a TIP for a 900 km diameter circle centred in north-central California, with a 100-km extension north and south, for an earthquake greater than magnitude 7.5, in effect until the end of 1991; second, a TIP for northern California, for an earthquake greater than magnitude 6.4, in effect until February 1989 (no event occurred, but it is not known if subsequent analysis extended this TIP).

NEPEC found the results of the seismicity analysis interesting and the conclusions concerning California intriguing. Remaining uncertainties about both the

approach and the conclusions are frankly addressed. The needs for further research are emphasized, including continued US-Soviet joint research. One recommendation is that the US programme should take more seriously the idea that the preparation zones for large earthquakes extend over large areas. Another recommendation calls for the "cleaning up" of existing regional and global earthquake catalogues and the development of new and improved ones. Routinely produced catalogues are being used as data sources for prediction in ways that go beyond the standard norms of completeness, accuracy and uniformity of data treatment.

NEPEC's bottom line on the Soviet prediction is "that the results do not at this time warrant any special public-policy actions in California and western Nevada". NEPEC members remain concerned

#### IMMUNODEFICIENCY VIRUSES

## The simian-human connection

Russell F. Doolittle

THE current AIDS pandemic is well known to have two foci: one in central Africa, associated with the immunodeficiency virus denoted HIV-1, and a second in West Africa, thought to be caused by a second population of retroviruses termed HIV-2. Where did these retroviruses come from? The paper by Hirsch *et al.* on page 389 of this issue<sup>1</sup> identifies the probable origin of the HIV-2 epidemic, and may help to resolve the controversy about the history of the spread of HIV-1.

From the outset — even before it was realized that there are two significantly different populations of HIV — attention centred on non-human primates as a possible source of cross-infecting virus. Large-scale serology surveys<sup>2-5</sup> on non-human primates, both on captive animals housed in primate colonies around the world and on freshly taken specimens in Africa, have been undertaken in search of animals with antibodies against HIV or closely related viruses. Indeed, seropositive individuals were found in several groups of Old World monkeys, including African green monkeys, mandrills and sooty mangabeys.

When the sequence of HIV-2 was reported<sup>6</sup>, it was clearly significantly different from HIV-1, although the two viruses were nevertheless more closely related to each other than to any other retroviruses whose sequences were known at the time. Shortly thereafter, the sequence of a virus isolated from a stricken captive macaque was determined<sup>7,8</sup> and found to be remarkably similar to HIV-2; it is now denoted SIV<sub>mac</sub>. The mystery was to an extent deepened by

about the high probability of strong and damaging earthquakes in the region in the next few decades, but conclude that "existing US analyses . . . constitute a suitable basis for the continuing development of public policy planning and actions". □

Carl Kisslinger is Professor of Geological Sciences at the University of Colorado at Boulder, Campus Box 216, Boulder, Colorado 80309, USA.

1. Wyss, M. & Burford, R.O. *Nature* **329**, 323–325 (1987).
2. Kisslinger, C. *Bull. seismol. Soc. Am.* **78**, 218–229 (1988).
3. Keilis-Borok, V.I., Knopoff, L., Rotwain, I.M. & Allen, C.R. *Nature* **335**, 690–694 (1988).
4. Udike, R.G. *US Geological Survey Open-File Report*, 89–144 (1989).
5. Wallace, R.E., Davis, J.F. & McNally, K.C. *Bull. seismol. Soc. Am.* **74**, 1819–1825 (1984).
6. Aki, K. & Stuart, W.D. *US Geological Survey Open-File Report*, 87–591 (1987).
7. *Earthquakes and Volcanoes* **20** No. 2 (US Geological Survey, 1988).
8. Working Group on California Earthquake Probabilities *US Geological Survey Open File Report*, 88–398 (1988).

these data, because the natural habitat of macaque monkeys is Asia and not Africa and, as far as could be told from the serology surveys, macaques in the wild did not harbour the virus. It was thought likely that the ailing macaques had contracted it from some other primate, probably mangabeys, during captivity<sup>9</sup>.

The paper by Hirsch *et al.* in this issue<sup>1</sup> provides strong evidence that this is indeed the case. The more important corollary of this study is that a sooty mangabey retrovirus is now the leading candidate for sparking the HIV-2 epidemic. Sooty mangabeys, unlike macaques, are indigenous to West Africa, the general locale of the human HIV-2 outbreak, and the virus is known to occur among mangabeys in the wild. Hirsch *et al.* make a good case for the fateful horizontal infection of a human occurring 30–40 years ago.

These findings hardly solve all the mysteries. In particular, where did HIV-1 come from? Consider the two reports that appeared in two consecutive issues of *Nature* last year. In one, Hayami's group in Japan reported<sup>10</sup> the sequence of a retrovirus isolated from naturally infected African green monkeys of Kenyan origin. The sequence of this virus, SIV<sub>agm</sub>, indicated that it was about equally distant from HIV-1 and HIV-2 (see figure); it appeared to be present in about 25 per cent of the green monkeys tested. Adopting the same stance they had taken previously in their studies of simian and primate T-cell leukaemia viruses, Hayami and co-workers claimed that immunodeficiency viruses are species-specific and that interspecific transmission in recent times

between monkeys and man is unlikely. (The title of a News and Views article<sup>11</sup> accompanying that paper was "Human AIDS Virus Not From Monkeys".) Hayami and co-workers believe that HIV-1 and SIV<sub>agm</sub> have diverged gradually with the divergence of men and monkeys. HIV-1 and HIV-2, they say, must have diverged among humans a long time ago and have been maintained in different isolated populations and in separated localities<sup>10</sup>.

In the sharply contrasting view that appeared in the next issue<sup>12</sup>, Smith *et al.* reported that HIV-1 and HIV-2 could have diverged as recently as 40 years ago. These authors based their estimate on the amount of change observed in individual HIV-1 and HIV-2 isolates and by extrapolating the results backwards in time. The data of Hirsch *et al.*<sup>1</sup> allow a similar interpretation. They reason that if SIV<sub>mac</sub> is in fact the result of a horizontal infection by a retrovirus from a sooty mangabey within the past 20–30 years — the extent of co-housing of the mangabeys and macaques — then the time back to other common ancestors can be gauged from the degree of difference between SIV<sub>sm</sub> and SIV<sub>mac</sub>. In agreement with other estimates<sup>3,14</sup> the extension of these data to HIV-1 and HIV-2 or HIV-1 and SIV<sub>agm</sub> results in divergence times of millions of years at most, and not of tens of millions of years as Hayami's group believes.

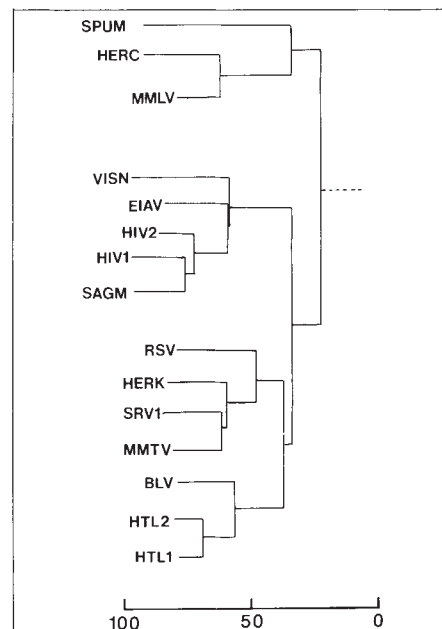
There are some remarkable aspects to this story. Quite apart from the astonishing ease with which horizontal retrovirus infection seems to occur among non-human primates housed together (there are numerous other examples, including the notorious gibbon-woolly monkey transfer<sup>15</sup> that gave rise to the oncogenic simian sarcoma virus), the most fascinating aspect of these heterologous involvements is that the primary hosts all seem to be healthy. The mangabeys that furnished the virus for the study of Hirsch *et al.*, for example, were all perfectly robust. When the virus was injected into macaque monkeys, however, the animals quickly became severely ill, ultimately dying of opportunistic infections. Similarly, the virus isolated from green monkeys<sup>10</sup>, SIV<sub>agm</sub>, has no effect on green monkey cells in culture, but human cells are rapidly attacked and show signs of cytopathic effects such as giant cell formation<sup>16</sup>.

Hayami's group might argue, as in the similar case of the T-cell leukaemia viruses and adult T-cell leukaemia<sup>17</sup>, that the primary host has had a very long time to adapt to the virus. This could explain why African green monkeys are apparently unaffected by SIV<sub>agm</sub>, but would argue against the contention that the HIVs have long been present in humans.

There is another aspect that is often ignored in discussing these issues: it seems impossible for a retrovirus to remain exogenous and infectious for millions of

years and yet retain a recognizable sequence. The rate of sequence change while the virus is under the replicative influence of its reverse transcriptase is overwhelming, and it is this rate that investigators are observing among the various HIV isolates. What must happen is that infectious retroviruses eventually invade the host germ line and become 'endogenized'. Once under control of the host replicative apparatus, the rate of retrovirus sequence change must be slowed by several orders of magnitude<sup>18</sup>.

Thus, much of the mystery can be dispelled by distinguishing the lifestyles of



Phylogeny of vertebrate retroviruses based on reverse transcriptase sequences<sup>23</sup>. HTL1, human T-cell leukaemia virus-I; HTL2, human T-cell leukaemia virus-II; BLV, bovine leukaemia virus; SRV1, simian retrovirus-1; MMTV, mouse mammary tumour virus; HERK, human endogenous retrovirus; RSV, Rous sarcoma virus; SAGM, SIV<sub>agm</sub>; HIV1, human immunodeficiency virus-1; HIV2, human immunodeficiency virus-2; EIAV, equine infectious anaemia virus; VISN, visna lentivirus; MMLV, mouse Moloney leukaemia virus; HERC, human endogenous retrovirus; SPUM, human spumavirus. For a closer view of HIV-2 relationships, see the tree by Hirsch *et al.* on page 391 of this issue.

exogenous and endogenous retroviruses. Exogenous (infectious) retroviruses are usually pathogenic, whereas endogenous retroviruses are carried in the germ line of a species (or a sub-population of a species), are usually benign and frequently defective. They often exist in multiple copies distributed about the genome, and in some cases it has been shown that these multiple copies have been frozen in place for millions of years. As a case in point, an endogenous retrovirus has been found to occur at the same multiple places in the genomes of humans

and chimpanzees, indicating that the genomic diaspora occurred before the divergence of these species<sup>19</sup>. Thus, in the phylogeny depicted in the figure, endogenous retroviruses that are millions of years old are interspersed with exogenous retroviruses such as HIV-1 and HIV-2 that have had separate existences marked by decades or centuries.

There is another feature of endogenous retroviruses that bears on the problem: many of them are able to infect species other than their original host<sup>20</sup>. Indeed, 'xenotropism' is defined as the resistance of animal cells to infection by their own endogenous viruses while related species remain susceptible<sup>22</sup>. Thus, some primate species may be refractory to the retroviruses being isolated from them because they are carrying the same or very similar viruses in their germ line.

Seen in this light, what is now the simplest interpretation of the immunodeficiency virus sequence data? Current betting would have to favour two separate horizontal infections of humans during the past 20–40 years. In one, an individual in West Africa was exposed to a retrovirus from a non-human primate, probably a sooty mangabey. In the other, someone in central Africa was infected by another retrovirus, perhaps from an African green monkey, but possibly from an as-yet unidentified primate. (HIV-positive mandrills have been reported<sup>5,16</sup>, for example, and a mysterious chimpanzee sample is being studied<sup>21</sup>.) It is not impossible, either, that a nonprimate carried the progenitor virus. This interpretation precludes HIV-1 and HIV-2 having shared a common ancestor in humans. How long the different offending retroviruses may have been resident in their wild hosts remains an open question. Do some sooty mangabeys carry SIV<sub>sm</sub> in their germ line? □

Russell F. Doolittle is a professor in the Center for Molecular Genetics, University of California, San Diego, La Jolla, California 92093, USA.

- Hirsch, V.M. *et al.* *Nature* **339**, 389–392 (1989).
- Barin, F. *et al.* *Lancet* **II**, 1387–1389 (1985).
- Kanki, P.J. *et al.* *Lancet* **I**, 1330–1332 (1985).
- Hayami, M. *et al.* *Jap. J. exp. Med.* **55**, 251–255 (1985).
- Lowenstine, L.J. *et al.* *Int. J. Cancer* **38**, 563 (1986).
- Guyader, M. *et al.* *Nature* **326**, 662–669 (1987).
- Franchini, G. *et al.* *Nature* **328**, 539–543 (1987).
- Chakrabarti, L. *et al.* *Nature* **328**, 543–547 (1987).
- Schneider, J. & Hunsman, G. *AIDS* **2**, 1–9 (1988).
- Fukasawa, M. *et al.* *Nature* **333**, 457–461 (1988).
- Mulder, C. *Nature* **333**, 396 (1988).
- Smith, T.F. *et al.* *Nature* **333**, 573–575 (1988).
- Yokoyama, S., Chung, L. & Gojobori, T. *Molec. Biol. Evol.* **5**, 237–251 (1988).
- Sharp, P.M. & Li, W.-H. *Nature* **336**, 315 (1988).
- Thielen, G.J. *et al.* *J. natn. Cancer Inst.* **47**, 881 (1971).
- Ohta, Y. *et al.* *Int. J. Cancer* **41**, 115–122 (1988).
- Hayami, M. *Cancer Rev.* **1**, 35–63 (1986).
- Doolittle, R.F., Feng, D.-F., Johnson, M.S. & McClure, M.A. *Q. Rev. Biol.* **64**, 1–30 (1989).
- Steele, P.E. *et al.* *J. Virol.* **59**, 545–550 (1986).
- Todaro, G.J. *Am. J. Path.* **81**, 590–605 (1975).
- New Scientist*, (9 June 1988).
- Coffin, J. in *RNA Tumor Viruses* (R. Weiss *et al.* eds) 2nd edn (Cold Spring Harbor, New York, 1984).
- Doolittle, R.F., Feng, D.-F., McClure, M.A. & Johnson, M.S. in *Current Topics in Microbiology and Immunology* (eds Vogt, P.K. & Swanstrom, R.) (Springer, in the press).



# Colloids stick to fractal rules

John Rarity

A TWO-DIMENSIONAL projection of a tree's root system looks remarkably like an overhead view of a river delta. Both are random fractals and the similarities arise from similar large-scale rules of growth which seem not to be influenced by their detailed constituents. The attraction of fractal theories of nature is that the application of such large-scale geometrical rules can lead to simple universal properties in many complex, random-growth processes. One such process that has been popular with the colloid scientist since the days of the Faraday gold sol is the phenomenon of colloidal aggregation. In their letter elsewhere in this issue (*Nature* **339**, 360–362; 1989), M.Y. Lin *et al.* present electron micrographs of aggregates of colloidal gold, polystyrene and silica: the obvious similarities between these provide compelling visual evidence for universalities in their reactions.

Colloids, being suspensions of small particles in a fluid, are thermodynamically unstable to aggregation and survive only because of strong repulsive forces between the particles. Aggregation is thus inevitable when repulsive barriers are reduced, by various means specific to the constituents of the colloid. Two rate-limiting processes control the growth of random aggregates from the suspension. The first is the time taken for clusters to diffuse into contact with each other to allow sticking to occur; the second is the probability of sticking on contact.

The large-scale rule which leads to fractal properties of the aggregates is simply the geometrical limitation on close diffusive approach of two irregularly shaped objects. If particles stick on first contact and the bonds are stiff, one expects quite tenuous structures to evolve. If there is a low sticking probability per contact, the aggregates will have time to explore the contact surface and can interpenetrate somewhat before sticking, leading to less tenuous structures. The fractal nature of the resulting structure arises because each successful collision leads to a drop in the mean density of the aggregate and, on average, we can relate the mass contained in the aggregate  $M$  to a measure of its radius  $R$  by a power law,  $M \propto R^d$ , where  $d$  is the fractal dimension, which can range from 3 for space filling objects (or aggregates whose density does not drop with mass) to 1 in the case of extremely tenuous (rod like) structures.

The two limiting regimes investigated by Lin *et al.* are described as diffusion-limited cluster aggregation (DLCA), in which aggregates stick on contact, and slow reaction-limited aggregation (RLCA) in which the sticking probability on contact is

extremely low. Following the simplistic argument above, one expects DLCA aggregates to be more tenuous than RLCA aggregates. This is clearly the case for electron micrographs presented by Lin *et al.* on page 360. The fractal dimensions of the aggregates can be estimated by eye by noting that as these are two-dimensional projections, objects with fractal dimension smaller than 2 appear semi-transparent (as do DLCA aggregates) whereas objects with fractal dimension greater than 2 appear solid (RLCA aggregates).

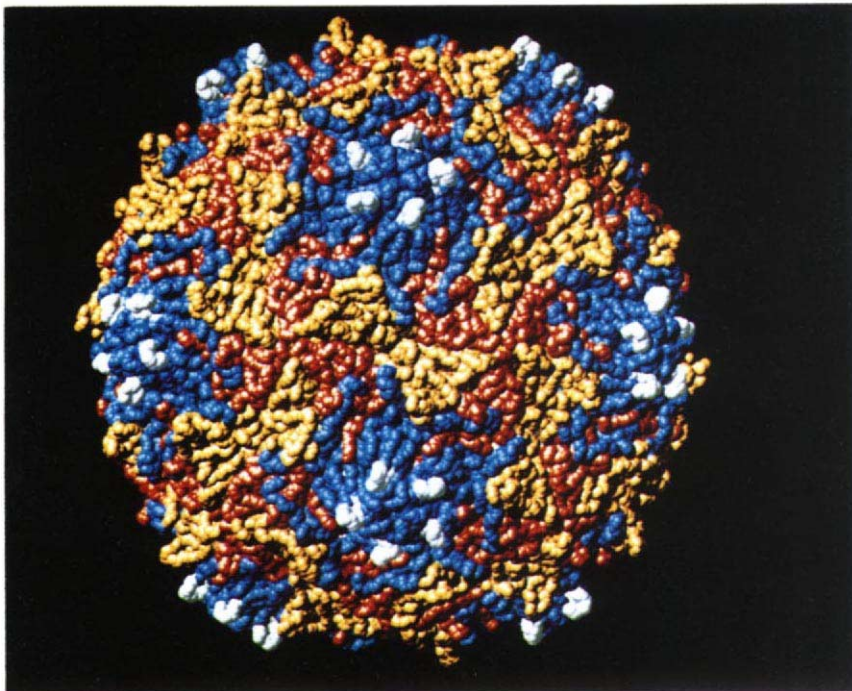
Lin *et al.* measure the fractal dimension of large aggregates in suspension using static (elastic) light scattering. The inverse of the scattering vector  $q$ , the difference between the input and scattered wave-vectors, defines a length scale (for coherent scattering) and the light scattered from an aggregate depends on the amount of mass typically found within a box of these dimensions. As a result the scattered intensity  $I(q)$  falls with increasing scattering angle following the power law form  $I(q) \propto q^{-d}$ . All three aggregated systems show fractal dimensions  $d \approx 2.1$  for RLCA and  $d \approx 1.85$  for DLCA as suggested by the micrographs.

There is also good reason to believe that

the kinetics of the aggregation reaction is also universal. The present workers follow the kinetics using dynamic light-scattering measurements to estimate a mean aggregate radius  $\bar{R}$  and show data which illustrate that  $\bar{R} \propto t^{1/d}$  (implying linear growth of aggregate mean mass) for DLCA and exponential growth,  $\bar{R} \propto e^t$ , for RLCA with all three systems studied. These data support the predictions of recent theory and simulations based on the simple rules of diffusion time and sticking probability governing the kinetics.

Is that it? Is the aggregation reaction now well understood? The neatly wrapped results presented here contrast sharply with those of new work by J.P. Wilcoxon *et al.* (*Phys. Rev. A* **39**, 2675–2688; 1989), who have been studying the aggregation of colloidal gold. Their article also reprints several earlier results from Lin *et al.* which at times disagree with those of the present work. Two main points of disagreement with the postulated universal kinetics are reported. First, Wilcoxon *et al.* point out that the strong optical resonance in aggregated gold at 680 nm can lead to measured fractal dimensions being wavelength dependent and report measurements of fractal dimensions ranging from 1.6 to 2 apparently uncorrelated with the aggregation kinetics. Second, they report a power-law growth of aggregate radius in DLCA with  $\bar{R} \propto t^{0.4}$  when from the above we expect an exponent of

## HIV/Poliovirus chimaera in AIDS research



THE outer surface of poliovirus, highlighting (in white) that part of the capsid protein, VP1, which has been replaced by part of the envelope glycoprotein of human immunodeficiency virus-1 in the chimaeric virus whose immunological properties are described by Evans *et al.* on page 385. The bulk of VP1 is shown in blue, whereas VP2 and VP3 are in yellow and red, respectively. (The figure was kindly prepared by C. Lockshin, D. J. Filman and J. M. Hogle of the Research Institute of Scripps Clinic.) □



$1/d \approx 0.55$ . Both these results do not necessarily contradict the results of Lin *et al.* but do illustrate the difficulty of unambiguously interpreting light-scattering data.

Lin *et al.* made all their new measurements at a wavelength of 488 nm, well away from the optical resonance and thus should lead to a reasonable estimate of fractal dimension. In earlier work, lower values of the DCLA fractal dimension were reported which may be attributable to neglect of optical resonance and multiple scattering effects. Dynamic light-scattering data can be ambiguous because aggregating systems contain broad size distributions of randomly shaped aggregates. For these systems the technique measures an apparent radius dependent on scattering vector, because rotational motion of larger aggregates distorts the

result. Lin *et al.* use a scaling approach based on multi-angle dynamic light scattering to estimate a true radius whereas Wilcoxon *et al.* use a simpler low-angle approach which may lead to distortion (in fact their data agree with uncorrected data from Lin's group).

The paper by Wilcoxon *et al.* does, however, highlight the earlier inconsistencies of Lin *et al.*, a typical problem of rushing into print in a rapidly moving field. Although Lin's group may need a little chiding for having in the past published interpretations and results that had flaws, the present work convinces me of universality in colloid aggregation. But there are exceptions to every rule....  $\square$

John Rarity is at the Royal Signals and Radar Establishment, St Andrews Road, Malvern, WR14 3PS, UK.

## MATHEMATICS

# Mock theta conjectures

Ian Stewart

ONE of the most unusual people in the annals of mathematical research is Srinivasa Ramanujan, a self-taught Indian mathematician whose premature death left a rich legacy of unproved theorems. Ramanujan was preeminent in an unfashionable field — the manipulation of formulas. He tended to state his results without proofs — indeed on many occasions it is unclear whether he possessed proofs in the accepted sense — yet he had an uncanny knack of penetrating to the heart of the matter. Over the years, many of Ramanujan's claims have been established in full rigour, although seldom easily. The most recent example, the 'mock theta conjectures', is especially striking, because the results in question were stated in Ramanujan's final correspondence with his collaborator Godfrey H. Hardy. The conjectures have recently been proved by Dean Hickerson (*Invent. Math.* **94**, 639–660; 1988), following fundamental work of George Andrews and F.G. Garvan (*Adv. Maths* **73**, 242–255; 1989).

The word 'theta' refers to a remarkable range of special functions, known as theta functions, investigated at great length by Carl Gustav Jacob Jacobi (1804–51). They arose during the nineteenth century and rapidly became a major research area. Attempts to calculate the length of a general arc of an ellipse using integral calculus lead to simple-looking expressions which obstinately refuse to yield an explicit integral in terms of known functions such as polynomials or trigonometric functions. The reason turns out to be that these integrals require a genuinely new breed of function, called 'elliptic' functions. In a sense, elliptic functions are generalizations

of the trigonometric functions, because an ellipse is a generalized circle and arc-lengths of circles involve trigonometry. But they have a deeper mathematical property. Trigonometric functions are periodic, that is, they repeat their values if a constant, the period, is added to the independent variable. The period of the sine function, for example, is  $2\pi$ , because  $\sin(x+2\pi) = \sin x$ . Elliptic functions are doubly periodic: they have two independent periods, which in general are complex numbers, not real ones. This property singles them out and renders them worthy of serious study as pure mathematics, whereas their many applications to problems in number theory and dynamics make them useful enough to earn their keep in a competitive entrepreneurial environment. Although they have all but disappeared from the undergraduate mathematics course, they remain important at the frontiers of research.

Jacobi introduced four associated functions, the theta functions, which in the words of Morris Kline (*Mathematical Thought from Ancient to Modern Times*; Oxford University Press, 1972) are "the simplest elements out of which the elliptic functions can be constructed". He also obtained expressions for them in the form of infinite series and infinite products and proved a number of remarkable identities.

Ramanujan's last letter to Hardy (S. Ramanujan in *Collected Papers*, 354–355; Cambridge University Press, 1927) introduces ten new functions, occurring in two groups of five. They are defined by series which in some respects are similar to Jacobi's, and Ramanujan asserted that they share many analogous properties, though, as usual, he gave no proofs. They

are accordingly known as mock theta functions and said to be of the 'fifth order' to distinguish them from related functions found elsewhere in Ramanujan's works.

But are they genuine mock theta functions, rather than ordinary Jacobi theta functions in some as yet unpenetrated disguise? Real fakes rather than fake fakes? It is a subtle question, but an important one. If Ramanujan's mock thetas are really real thetas, they immediately lose any interest. Of course fancy combinations of genuine thetas behave like genuine thetas. As with certain art forgeries where the original artist has gone out of fashion but the forger has acquired notoriety, only a genuine fake is interesting. To put it another way: real thetas are interesting, and mock thetas share their remarkable properties, so provided that they are really new, mock thetas are also interesting.

And that is where the mock theta conjectures come in. Ramanujan recorded his results in a series of notebooks. One temporarily went missing, becoming the romantic 'lost notebook', but eventually turned up safe and sound. Following early work of G.N. Watson, George Andrews discovered a formula in the lost notebook which, if true, means that the mock thetas are genuine fakes. In his paper with Garvan cited above, he shows that this formula is equivalent to two slightly involved statements in number theory. These concern the partitions of a given integer — the ways of writing it as a sum of smaller integers. The original papers should be consulted for the details.

Hickerson has now proved these number-theoretic conjectures, so the mock thetas are genuine fakes and are therefore interesting new functions worthy of further study. The proof involves delicate manipulations of infinite series of a kind that would have delighted Ramanujan. The astonishing complexity of the proof underlines, yet again, the depth of Ramanujan's genius. It is very hard to see how anyone could have been led to such results without getting bogged down in the fine detail. Hickerson has extended his methods (*Intent. Math.* **94**, 661–677; 1988) to tackle the seventh order mock theta functions, also introduced by Ramanujan and equally enigmatic. They too prove to be genuinely mock.

Ramanujan was the formula man *par excellence*, operating in a period when formulas were out of fashion. Today's renewed emphasis on combinatorics, inspired in part by the digital nature of computers, has provoked a renewed interest in formal manipulations. The half-forgotten ideas of Srinivasa Ramanujan are breathing new life into number theory and combinatorics.  $\square$

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

# Antigens for $\gamma/\delta$ T cells

David H. Raulet

FOR some time after the T-cell receptor  $\gamma$ -chain gene was discovered five years ago<sup>1</sup> it was considered by many immunologists to represent a curiosity at best. Most attention at that time focused on understanding the workings of the newly discovered  $\alpha\beta$  T-cell receptor heterodimer, which is expressed by most T cells. Widespread interest in the  $\gamma$ -chain was piqued, however, by the subsequent discovery that  $\gamma$  and another chain,  $\delta$ , compose a second T-cell receptor heterodimer, expressed by a separate subtype of T lymphocytes (called  $\gamma/\delta$  T cells). But answers to important questions about  $\gamma/\delta$  T cells — what do they recognize and why? — have remained elusive. The first reports of  $\gamma/\delta$  T cells that recognize conventional antigens, those of various mycobacteria, are just beginning to appear<sup>2-5</sup>. Moreover, the data imply that at least a significant fraction of  $\gamma/\delta$  cells are specialized to recognize epitopes on a class of phylogenetically conserved stress or 'heat-shock' proteins which have previously been implicated as immunodominant mycobacterial antigens.

## Mycobacterial immunity

The obvious conclusion is that  $\gamma/\delta$  T cells are particularly specialized for mycobacterial immunity. But this interpretation must be tempered somewhat by various observations about stress proteins, about  $\gamma/\delta$  cells in general, and about some of the mycobacteria-reactive T cells that have just been characterized. A competing, although not mutually exclusive, hypothesis is that at least a subset of  $\gamma/\delta$  T cells is devoted to the recognition of conserved epitopes on stress proteins, including autologous stress proteins, and may represent a primitive arm of immunity devoted to the elimination of an individual's transformed, infected or otherwise stressed autologous cells<sup>3,6,7</sup>.

Several different approaches have been used in the present studies to demonstrate mycobacteria-reactive  $\gamma/\delta$  T cells. Holoshitz *et al.*<sup>2</sup>, relying on past findings that T lymphocytes from synovial fluid of rheumatoid arthritis patients proliferate in response to mycobacterial antigens, used this approach to expand and clone human synovial T cells reactive to *Mycobacterium tuberculosis* antigens. Four of the five clones isolated turned out to be  $\gamma/\delta$  T cells of the CD4<sup>+</sup>CD8<sup>-</sup> phenotype and to proliferate in response to mycobacterial antigens. In addition, one of the  $\gamma/\delta$  clones (the only one tested) responds to a purified preparation of the HSP60 stress protein from *M. bovis* (known from DNA sequencing to be identical in sequence to the *M. tuberculosis* HSP60).

Using a similar approach, Modlin *et al.*<sup>5</sup> find that  $\gamma/\delta$  T cells are enriched in certain lesions of patients with leprosy or leishmaniasis. They derived a  $\gamma/\delta$  T-cell line reactive with *M. leprae* antigens from patients with leprosy, which proliferates *in vitro* in response to *M. leprae* cell-wall antigens as well as to the 'purified protein derivative' (PPD) — a relatively crude preparation of mycobacterial antigens — of *M. tuberculosis*. In contrast to the findings of Holoshitz *et al.*, Modlin *et al.* observe no response of these  $\gamma/\delta$  T cells to the *M. bovis* HSP60 stress protein. Janis *et al.*<sup>4</sup> immunized mice with PPD from *M. tuberculosis*, and showed that T cells from the draining lymph nodes are enriched for  $\gamma/\delta$  T cells of the CD4<sup>+</sup>CD8<sup>-</sup> phenotype. They then showed that  $\gamma/\delta$  T cells purified from the draining nodes respond specifically to PPD *in vitro*.

O'Brien *et al.*<sup>3</sup> used a different starting point altogether, and their results provide a different perspective to all these findings. Their study began with the analysis of a panel of murine  $\gamma/\delta$  hybrids isolated from an antigen-unselected population of neonatal thymocytes. While analysing the hybrids for alloreactivity (none was found), control experiments showed that more than a third of the cell lines 'spontaneously' produce interleukin-2 in the absence of any added antigens. Several of these authors' experiments suggest that the 'spontaneous' production of interleukin-2 requires the participation of the  $\gamma/\delta$  receptors on the cells, including the finding that almost all these hybrids are distinguished by the expression of members of a particular  $V_\delta$  gene family ( $V_\delta 6$ ). The authors propose that some  $\gamma/\delta$  receptors, particularly those using  $V_\delta 6$  regions, are autoreactive with an antigen expressed by the T hybrids themselves.

The relation of the results of O'Brien *et al.* to mycobacteria became apparent from further attempts to define the antigen specificities of the hybrids. Of several antigens tested, only *M. tuberculosis* PPD elicits responses from any of the hybrids; strikingly, the hybrids that respond are precisely those that also spontaneously respond without antigen (the response to PPD is an increase over the spontaneous response of these hybrids). These results both show that a large proportion of antigen-unselected  $\gamma/\delta$  cells from the newborn thymus respond to PPD, and demonstrate an exact correlation between PPD reactivity and 'spontaneous' interleukin-2 production. This means that PPD reactivity also correlates (imperfectly) with expression of  $V_\delta 6$  family members. Further testing of the PPD-reactive hybrids shows that most of them respond

weakly to recombinant *M. bovis* HSP60.

As the authors point out<sup>3</sup>, these findings raise the interesting possibility that this group of T hybrids reacts with autologous (murine) and mycobacterial stress proteins. Considering the remarkable sequence conservation between bacteria and mammals of some of these proteins (including HSP60, which is 50 per cent identical), this suggestion merits serious consideration, particularly in the light of a new study<sup>8</sup> showing that human T-cell lines raised against mycobacterial HSP60 peptides crossreact with the human heat-shock protein homologue.

Several features of  $\gamma/\delta$  T cells and their repertoire (reviewed in ref. 9) had earlier suggested the hypothesis that at least some of them are specialized for destruction of 'stressed' autologous cells<sup>6,7</sup>. First, although  $\gamma/\delta$  cells represent a minor fraction of lymphocytes in the secondary lymphoid organs, they are by far the main lymphocyte type resident in the epithelial layers of the epidermis and intestines in the mouse. Although there are variations on this theme depending on the species, the epithelial tropism of some  $\gamma/\delta$  T cells has led to the notion that these cells are devoted to monitoring the integrity of epithelial cell layers. Second, the expressed diversity of  $\gamma/\delta$  receptors in certain sites is highly restricted, despite the significant germline diversity of the  $\gamma/\delta$  T-cell receptor (there are seven  $V_\gamma$  genes and more than ten  $V_\delta$  genes in the mouse), and the potential for extreme junctional diversity, particularly in the  $\delta$ -chain. Most remarkable are the murine epidermal  $\gamma/\delta$  T cells, most of which seem to express identical receptor sequences, even in the junctional regions<sup>7</sup>. Because intraepithelial  $\gamma/\delta$  cells in the epidermis contact neighbouring keratinocytes, are largely immobile, and have a highly restricted repertoire, it was suggested that they recognize a common antigen expressed by stressed (infected or transformed) keratinocytes. In line with earlier observations that epidermal  $\gamma/\delta$  T cells can express cytolytic activity, it was proposed that  $\gamma/\delta$  cells function to eliminate stressed autologous cells.

## Recognition

On the other hand, some features of  $\gamma/\delta$  cells in other sites in the body seem more consistent with a role in immunity against foreign antigens. Most striking in this vein are recent studies<sup>10</sup> showing that a significant fraction of intestinal intraepithelial  $\gamma/\delta$  T cells in normal, but not germ-free, mice are activated *in situ* (that is, freshly isolated cells express cytolytic activity without *in vitro* stimulation). When the intestinal flora of germ-free mice is repopulated by removing the mice from barrier conditions, their intestinal  $\gamma/\delta$  cells become activated like those in normal mice<sup>10</sup>. These findings led to the



proposal that intestinal  $\gamma/\delta$  T cells function in immunity to intestinal pathogens. In line with this idea, a new study<sup>11</sup> reveals that  $\gamma/\delta$  receptors on intestinal intraepithelial T cells use several  $V_\delta$  regions and exhibit considerable junctional diversity of both  $\gamma$ - and  $\delta$ -chains.

So are (some)  $\gamma/\delta$  cells specialized for recognition of mycobacterial antigens, or is the reactivity with mycobacterial antigens a crossreactivity of cells specialized for recognition of autologous stress proteins? Perhaps both ideas are correct. The  $\gamma/\delta$  T cells may have originally evolved to eliminate stressed self-cells, and later evolved some degree of diversity to recognize foreign antigens<sup>9</sup>. According to this idea,  $\gamma/\delta$  cells represent the most primitive immune cells. Perhaps some  $\gamma/\delta$  cells can have a dual function, eliminating stressed autologous cells as well as providing immunity to mycobacteria.

It is becoming clear that stress proteins are immunodominant antigens for anti-mycobacterial T cells in general<sup>12</sup>. There is also the intriguing suggestion that responses to mycobacteria may be involved in the genesis of certain autoimmune diseases. T cells from synovial fluid of rheumatoid arthritis patients, for example, respond *in vitro* to *M. tuberculosis* antigens<sup>2</sup>. Furthermore, it has been shown in previous studies that immunization of rats with *M. tuberculosis* can lead to the development of arthritis, as can transfer to normal rats of T-cell clones specific for mycobacterial stress proteins (HSP60)<sup>13</sup>. These findings could be related to the results of Holoshitz *et al.*<sup>2</sup> demonstrating HSP60-reactive  $\gamma/\delta$  T cells among synovial cells of rheumatoid arthritis patients, and begs the question as to whether  $\gamma/\delta$  T cells are involved in autoimmunity. The idea is that T cells ( $\alpha/\beta$  and/or  $\gamma/\delta$ ) elicited by mycobacterial infections crossreact on autologous stress proteins, which may be expressed at high levels by stressed cells at sites of infection<sup>8,13</sup>.

It is known that  $\alpha/\beta$  T cells recognize antigens bound to polymorphic major histocompatibility complex (MHC) proteins, and distinguish allelic MHC proteins; this phenomenon is called MHC restriction. How do the new papers on  $\gamma/\delta$  cells relate to this phenomenon? No evidence for 'conventional' MHC-restriction is apparent in the study of Holoshitz *et al.*, as heterologous antigen-presenting cells stimulate the  $\gamma/\delta$  clones used in this study. In contrast, the T-cell lines of Modlin *et al.*<sup>3</sup> respond poorly or not at all to antigen presented by allogeneic cells; however, the limited analysis does not allow the identification of the genes restricting the response. Interestingly, the T hybrids of O'Brien *et al.*<sup>3</sup> respond to PPD in the absence of other cells. Although they do not have direct evidence, O'Brien *et al.* suggest that the putative restricting

element is present on the hybrids themselves. Because the hybrids do not express class II MHC proteins, the putative restricting element would have to be a class I protein or not an MHC protein at all. The idea that  $\gamma/\delta$  cells are restricted by class I proteins encoded in the Qa/T1a region neighbouring the murine MHC genes has been around for some time<sup>6,9,14,15</sup>, largely because  $\gamma/\delta$  cells specific for class I molecules encoded by genes in this region have been isolated<sup>14,15</sup>, and because some  $\gamma/\delta$  cells express the CD8 molecule. The idea that these class I-like proteins are restricting elements for  $\gamma/\delta$  cells is also appealing because of the low level of genetic polymorphism in these proteins, which might account for the apparent lack of MHC-restriction observed in some instances<sup>2</sup>. On the other hand,  $\gamma/\delta$  T cells recognizing classical MHC allo-antigens have also been isolated<sup>14</sup>.

Working out the genetic restrictions governing recognition of mycobacterial antigens by  $\gamma/\delta$  T cells clearly requires much more information. But with anti-

NEURONAL-GLIAL INTERACTIONS

## A new form of transmission?

Barbara A. Barres

NEURONS signal each other by releasing chemical transmitters, a process thought to be confined to synapses. Two studies reported by Marrero *et al.* on page 378 and by Usowicz *et al.* on page 380 of this issue<sup>1,2</sup> provide evidence, albeit indirect, for non-synaptic axonal transmitter release, and suggest that the recipient is a glial cell.

The first evidence that neurotransmitter release can occur from axon tracts was the observation 20 years ago of impulse-dependent release of glutamate from nerves preloaded with <sup>14</sup>C-glutamate<sup>3,4</sup>. Although these findings seem to have provoked little interest, they take on new significance with the report of Usowicz *et al.*<sup>2</sup> that type 2 astrocytes — a type of glial cell thought to occur mainly in the white matter of the mammalian central nervous system<sup>5</sup> — express glutamate receptors. The suggestion that astrocytes have glutamate-activated ion channels, which was first made because of the ability of glutamate to depolarize glial cells, has been controversial<sup>6,7</sup>. Because neurotransmitter uptake is a long revered function of glia, the debate seemed to end with the demonstration that glutamate-induced depolarization of both Muller cells and cerebellar type-1 astrocytes is caused by an electrogenic glutamate-uptake mechanism<sup>8,9</sup>. But the recent report by Sontheimer *et al.*<sup>10</sup> that glutamate activates an ionic conductance in cortical (type-1-like) astrocytes re-opened the debate. These data are intriguing, but the authors did not

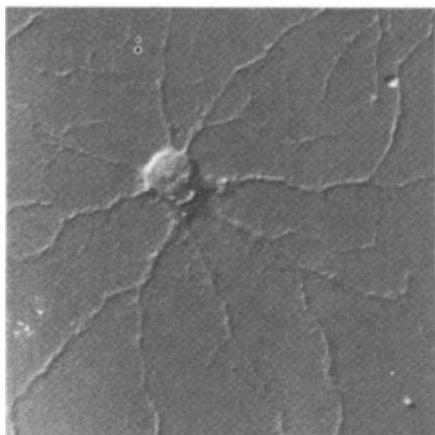
measure blockade of currents by glutamate antagonists, single-channel currents, or demonstrate that the glutamate-induced currents could be terminated by removing the transmitter from the bath.

Usowicz *et al.*<sup>2</sup> now report conclusive evidence for glutamate-activated ion channels in type 2 astrocytes in culture: they observe whole-cell and single-channel glutamate-activated currents that reverse near zero (as expected for an ion channel permeable to both sodium and potassium) and are largely blocked by the glutamate antagonist kynurenic acid. Two main pharmacological classes of ion-channel-coupled glutamate receptors are generally found together on neurons: NMDA receptors, thought to be important in long-term potentiation; and non-NMDA receptors, thought to mediate fast excitation. Type 2 astrocytes have only the non-NMDA receptors. Their associated ion channels have the same multiple conductance levels found in neurons, but there also seem to be specific differences. Of particular interest, Usowicz *et al.* observe that non-NMDA agonists routinely activate a large substate of 45 picosiemens in the type 2 astrocyte. This is unexpected because in neurons the 45-picosiemens substate is activated mainly by NMDA. It remains to be shown that these neurotransmitter receptors occur on glia *in vivo* as well as in culture, but this is likely to occur as it is now certain that there are other ion-channel types in glial cells *in vivo*<sup>11</sup>.

1. Saito, H. *et al.* *Nature* **309**, 757–762 (1984).
2. Holoshitz, J. *et al.* *Nature* **339**, 226–229 (1989).
3. O'Brien, R. *et al.* *Cell* **57**, 667–674 (1989).
4. Janis, E.M., Kaufman, S.H.E., Schwartz, R.H. & Pardoll, D.M. *Science* **244**, 713 (1989).
5. Modlin, R.L. *et al.* *Nature* (in the press).
6. Janeway, C.A., Jones, B. & Hayday, A. *Immun. Today* **9**, 65–95 (1988).
7. Asarnow, D.M. *et al.* *Cell* **55**, 837–847 (1988).
8. Lamb, J.R. *et al.* *Int. Immun.* (in the press).
9. Raulet, D.H. *A. Rev. Immun.* **7**, 175–207 (1989).
10. Lefrançois, L. & Goodman, T. *Science* **243**, 1716–1718 (1989).
11. Takagaki, Y., Decloux, A., Bonneville, M. & Tonegawa, S. *Nature* (in the press).
12. Young, D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 4267–4270 (1988).
13. van Eden, W. *et al.* *Nature* **331**, 171–173 (1988).
14. Bluestone, J.A. *et al.* *J. exp. Med.* **168**, 1899–1916 (1988).
15. Bonneville, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).

Although the biophysical results are of interest, the presence of glutamate-activated ion channels on type 2 astrocytes has a broader neurobiological significance. The type 2 astrocyte (see figure), first described by Raff and colleagues<sup>8,12</sup>, has been found only in white matter (whether an equivalent type exists in grey matter is still unknown). As can be seen from the figure, its stellate appearance in culture is somewhat neuronal, a resemblance further supported by its surface antigens<sup>5</sup> and ion channels<sup>13</sup>. *In vivo*, the processes of this cell are closely apposed to nodes to Ranvier<sup>14,15</sup>.

Thus if glutamate receptors on type 2 astrocytes are extrasomatic (as is the case in neurons), then type 2 astrocyte perinodal processes can easily sense any potential glutamate release at nodes during impulse activity. Such release could be mediated by a non-vesicular carrier mechanism as vesicle fusion has



A type 2 astrocyte in culture.

not been observed along axons or nodes. Although glia might also be the source of the released glutamate — because they have electrogenic glutamate uptake that could run in reverse<sup>8,9</sup> — this is unlikely as they also contain cytoplasmic enzymes that rapidly metabolize glutamate. (In either case, it remains to be shown that electrogenic glutamate-carrier mechanisms can mediate net glutamate release *in vivo*.)

Marrero *et al.*<sup>1</sup> suggest that axons release transmitters which can alter the properties of glial ion channels. They record from glial cells at the surface of the frog optic nerve using the loose patch-clamp technique, which allows voltage-clamp recording when gigaohm-seal patches cannot be obtained, and observe sodium currents, probably originating from glial cells (although an axonal contribution is not completely excluded). Axonal impulses alter these sodium currents by shifting their current-voltage relation in a hyperpolarizing direction. One possible explanation for the effect is the impulse-dependent release of an unidentified chemical factor from axons; more likely, however, is a change in local

ionic concentrations, as the authors suggest.

New evidence that glutamate mediates axon-to-Schwann-cell signalling in the squid giant axon<sup>16</sup> more strongly supports the idea of non-synaptic axonal release of neurotransmitter. These Schwann cells have non-NMDA glutamate receptors that are activated during nerve stimulation: impulses elicit a Schwann-cell depolarization that is blocked by specific glutamate antagonists. The main transmitter present in the axons of this invertebrate preparation is glutamate; its high concentration strongly implicates the axons as the source of the released glutamate.

Why would neurons want to activate glial channels during impulse activity? Usowicz *et al.* suggest that activation of glutamate receptors could alter local concentrations of sodium and potassium near the node, altering excitability. Alternatively, glutamate-induced depolarization could activate the voltage-dependent calcium channels also present in type 2 astrocytes<sup>13</sup>; this could in turn alter local calcium concentrations or affect release of glial neurotransmitters. Another hypothesis proposes impulse-dependent release of humoral substances from axons that activate glial chloride channels as part of a potassium homeostatic mechanism<sup>13</sup>. In any case, these novel glial channels suggest the existence of dynamic neuronal-glial signalling processes in parts of the nervous system long thought to be passive transmitters of information. Whether such signalling simply allows glial cells better to fulfill a role of servitude to neurons, or instead functions in information processing, is now a significant question. □

Barbara A. Barres is in the Program in Neuroscience, Harvard Medical School, Massachusetts General Hospital, Boston, Fruit Street, Massachusetts 02115, USA.

- Marrero, H., Astion, M. L., Coles, J. A. & Orkand, R. K. *Nature* **339**, 378–380 (1989).
- Usowicz, M. M., Gallo, V. & Cull-Candy, S. G. *Nature* **339**, 380–383 (1989).
- Wheeler, D. D., Boyarsky, L. L. & Brooks, W. H. *J. cell. Physiol.* **67**, 141–148 (1966).
- Weinreich, D. & Hammerschlag, R. *Brain Res.* **84**, 137–142 (1975).
- Raff, M. C., Abney, E. R., Cohen, J., Lindsay, R. & Noble, M. J. *Neurosci.* **3**, 1289–1300 (1983).
- Tang, C. & Orkand, R. K. *Neurosci. Lett.* **63**, 300–304 (1986).
- Kettenmann, H. & Schachner, M. *J. Neurosci.* **5**, 3295–3301 (1985).
- Brew, H. & Attwell, D. *Nature* **327**, 707–709 (1987).
- Cull-Candy, S. G., Howe, J. R. & Ogden, D. C. *J. Physiol., Lond.* **400**, 189–222 (1988).
- Sontheimer, H., Kettenmann, H., Backus, K. H. & Schachner, M. *Glia* **1**, 328–336 (1988).
- Barres, B. A., Chun, L. L. Y. & Corey, D. P. A. *Rev. Neurosci.* (in the press).
- Miller, R. H., ffrench-Constant, C. & Raff, M. C. *A. Rev. Neurosci.* **12**, 517–534 (1989).
- Barres, B. A., Chun, L. L. Y. & Corey, D. P. A. *Glia* **1**, 10–30 (1988).
- ffrench-Constant, C. & Raff, M. C. *Nature* **323**, 335–338 (1986).
- Miller, R. H., Fulton, B. P. & Raff, M. C. *Eur. J. Neurosci.* (in the press).
- Lieberman, E. M., Abbott, N. J. & Hassan, S. *Glia* **2**, 94–102 (1989).

## Plastic wood

WOOD is one of the fast-vanishing resources of the planet. The rain forests are being destroyed largely because of the industrial world's demand for timber. And yet much of this wood (small branches, offcuts and so on) is wasted, and much of the rest is used for transient structures like shuttering. Almost none of it is recycled. Daedalus contrasts this sad state of affairs with the active recycling of metal scrap. Unlike the metals, wood cannot be melted and re-cast into new bulk material.

But wood is not totally intractable. It can be softened and bent even by steam — hence such charming artefacts as the traditional bentwood chair. At higher temperatures, of course, it starts to decompose: an exothermic runaway process. Daedalus plans to control and limit this reaction by conducting it in pressurized water (whose critical temperature of 374 °C is well above the charring-point of wood). As the temperature rises, the wood should slowly soften and depolymerize to the point of melting, or at least of plasticization. Under the right conditions, this controlled hydrolysis and depolymerization should be almost perfectly reversible, like the thermal depolymerization of many plastics.

So DREADCO's engineers are submitting all sorts of wood samples to controlled cycles of heat and pressure in various aqueous media. Their ultimate goal is a process which will take in scrap wood, soften it to high plasticity, and then roll it out or extrude it into fresh baulks and planks of timber. Much of its grain and structure should survive, giving a product that is recognizably wood, and which still has the outstanding structural virtues of the original. Indeed, it may be even better — free of cracks and knot-holes and available in shapes and sizes impossible for any tree.

Thus the destructive profligacy of modern timber technology will be checked. Sawdust, demolition wood-waste, offcuts, shuttering and timber scrap of all kinds will all be wonderfully recycled into new good wood. Some scrap (like old desks and grand pianos) may need a melt-filtering step to remove nails and metal wire, but with good luck almost all the world's wood could be recovered for re-use.

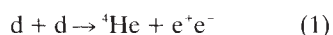
Even better, quite new methods of fabrication will open up. Complex items like wardrobes and bookcases could be moulded, in one operation, from plasticized wood. New materials like wire- or fibre-reinforced wood may become possible; even foamed wood, combining lightness and rigidity with thermal insulation. And subtle wood 'alloys', intimate mixtures combining the strength of spruce with the finish of mahogany, or the suppleness of bamboo with the durability of teak, should be possible.

David Jones

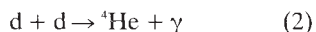


# Cold fusion results still unexplained

SIR—In the experiments which purport to demonstrate cold fusion of deuterons, neutron fluxes are low<sup>1</sup> but the heat output is high<sup>2</sup>, which has led to suggestions that deuterium fusion leading directly to <sup>4</sup>He somehow dominates over the more usual reactions. To consider a possible but improbable mechanism, I have calculated the rate of the reaction



in comparison with the reaction



on which extensive low-energy measurements have been made<sup>3</sup>. The rate of reaction (2) is known<sup>4</sup> to be about  $10^{-7}$  times that of the  $d+d \rightarrow n+{}^3\text{He}$  and  $d+d \rightarrow p+t$  reactions at temperatures below  $10^9$  K. I conclude that the rate of reaction (1) is lower still, by an additional factor of  $1.5 \times 10^{-2}$ , so cannot account for the claimed cold fusion.

Reaction (2) is a radiative-capture reaction, the theory of which has been treated in detail<sup>5</sup>. Resonance is not involved. Nevertheless, I prefer to express my results in terms of transition or  $\hbar$  times the transition probability per second. I find the  $\gamma$ -ray width in reaction (2) to be<sup>6</sup>

$$\Gamma_\gamma = \frac{\pi}{675} \alpha Z^2 \frac{E_\gamma^5}{(\hbar c)^4} a_\gamma^4 f_\gamma \quad (3)$$

where  $\alpha$  is the fine-structure constant,  $E_\gamma = 23.847$  MeV is the  $\gamma$ -ray energy (neglecting kinetic energy of the deuterons at low temperature),  $a_\gamma$  is a suitably defined radius for quadrupole transitions,  ${}^5S_2 \rightarrow {}^5D_0$ , of the electronic states (see equations (3) and (4) in ref. 6) and  $f_\gamma = 0.068$  is the  $D$ -state admixture<sup>6</sup> in the ground state of <sup>4</sup>He. Equation (3) is consistent with results obtained by others<sup>5,7</sup>.

Reaction (1) can be considered to be a capture reaction similar to reaction (2) but with different products. I find the pair-emission ( $\pi$ ) width in reaction (1) to be<sup>8</sup>

$$\Gamma_\pi = \frac{1}{375\pi} \alpha^2 Z^2 \frac{E_\pi^5}{\hbar c^4} a_\pi^4 f_\pi \quad (4)$$

where  $E_\pi = E_\gamma - 2m_e c^2 = E_\gamma - 1.022$  MeV is the pair-emission kinetic energy,  $a_\pi$  is a suitably defined radius for electric monopole transitions,  ${}^1S_0 \rightarrow {}^1S_0$ , and  $f_\pi = 0.932$  is the  $S$ -state admixture<sup>6</sup> in the ground state of <sup>4</sup>He.

The ratio of equations (4) and (3) yields

$$\frac{\Gamma_\pi}{\Gamma_\gamma} = \frac{9}{5\pi^2} \alpha \left( \frac{f_\pi}{f_\gamma} \right) (1 - 2m_e c^2/E_\gamma)^5 = 1.465 \times 10^{-2} \quad (5)$$

In computing this ratio I have set  $a_\pi = a_\gamma$  and I estimate an uncertainty of no more than 20 per cent in this equality. Thus

equation (5) holds to within a factor of about 2 and reaction (1) as well as all other  $d + d$  reactions cannot be the source of the claimed cold fusion of deuterons.

My conclusion is reinforced by experimental measurements<sup>9</sup> in <sup>16</sup>O; these indicate that the  $0^+$  pair-emitting state at 6.0494 MeV has a mean lifetime  $\tau_m = 96 \pm 7$  ps, whereas the  $2^+$  quadrupole  $\gamma$ -ray-emitting state at 6.9171 MeV has a mean lifetime  $\tau_m = 6.78 \pm 0.19$  fs. The inverse ratio  $\tau_m(2^+)/\tau_m(0^+) = \Gamma_\pi/\Gamma_\gamma = 7.06 \times 10^{-5}$ . In this case equation (5) can be used with  $f_\pi = f_\gamma$  and the last term in brackets replaced by  $(5.0274/6.9171)^5$ . The result is  $\Gamma_\pi/\Gamma_\gamma = 2.70 \times 10^{-4}$ , which is high by a factor of 3.8. However, in this case it is reasonable to expect  $a_\pi > a_\gamma$ , because of the higher angular momentum of the  $\gamma$ -ray-emitting state, and a ratio  $a_\pi/a_\gamma = 1.4$  removes the discrepancy with experiment.

WILLIAM A. FOWLER

W.K. Kellogg Radiation Laboratory,  
California Institute of Technology,  
Pasadena, California 91125, USA

1. Jones, S.E. et al. *Nature* **338**, 737–740 (1989).
2. Fleischmann, M. & Pons, S.J. *electroanal. Chem.* **261**, 301–308 (1989).
3. Barnes, C.A., et al. *Phys. Lett. B* **197**, 315 (1987).
4. Caughlan, G.R. & Fowler, W.A. *Atomic Data and Nuclear Data Tables* **40**, 283–334 (1988).
5. Christy, R.F. & Duck, I. *Nucl. Phys.* **24**, 89–101 (1961).
6. Assenbaum, H.-J. & Langanke, K. *Phys. Rev. C* **36**, 17 (1987).
7. Moszkowski, S.A. *Beta- and Gamma-Ray Spectroscopy* (ed. Siegbahn, K.) 373–395 (North-Holland, Amsterdam, 1955).
8. Oppenheimer, J.R. & Schwinger, J.S. *Phys. Rev.* **56**, 1066–1067 (1939).
9. Aizenberg-Selove, F. *Nucl. Phys.* **A460**, 22 (1986).

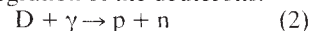
SIR—Excessive energy release has been cited as part of the evidence for cold fusion of deuterium dissolved in palladium<sup>1,2</sup>. I propose that the energy release, although nuclear in origin, is not due to fusion of deuterium but rather to a chain reaction involving radiative capture, by palladium nuclei, of neutrons produced by photodisintegration of deuterons.

Consider as a model situation an atomic pile comprising thin sheets of metallic palladium embedded in a moderator of solid deuterium. What would be the fate of a (possibly energetic) neutron in such an environment? Apart from elastic collisions in the moderator, the most likely event would be radiative capture by Pd nuclei; for example

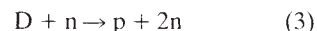


where the  $\gamma$ -ray energy is 9.95 MeV, a typical neutron-binding energy. The cross-section of this reaction, for neutron energies in the MeV range, is typically  $\sim 100$  millibarns, corresponding to a mean free path of the order of 10 cm in the metal. At lower energies the mean cross-section averaged over resonances is of the order of tens of barns (ref. 3). What happens to the  $\gamma$ -ray? As its energy is well above the relevant threshold (2.22 MeV), a signifi-

cant reaction in this context would be photodisintegration of the deuterons:



followed by neutron-induced disintegration:



for which sequence the total threshold energy is 5.55 MeV. Clearly in reactions (2) and (3) there are the makings of a chain reaction, in which each liberated neutron creates several more.

The cross-section for reaction (2) is of the order of several millibarns, and that for (3) is of the order of hundreds of millibarns<sup>4,5</sup>. The corresponding mean free path would be of the order of several metres at normal densities, which would be the required scale of the assembly if the reactions are to be self-sustaining. Elastic (Compton) scattering of electrons would compete with reaction (2), particularly as its cross-section at these photon energies is an order of magnitude higher than that for (2). This is not to say that the photon energy would be lost; energetic electrons so created would regenerate energetic photons via *bremssstrahlung*, resulting in a photon/electron cascade. A detailed mathematical model of this degradation process would be required to settle the question of net neutron yield on the basis of reactions (2) and (3); one suspects that they would fail to be self-sustaining, but not by a large factor.

My highly speculative scheme was motivated by the recent reports<sup>1,2</sup> of cold fusion in electrochemical cells comprising elements similar to those described above. I suggest that conditions within the palladium lattice, for example enhanced deuteron density, might be sufficient to initiate the chain reaction. An attractive feature of the proposed scheme, in comparison with fusion proper, is that of energy generation without commensurate neutron emission, which is what seems to be observed; energy is generated here as neutrons are captured, rather than released. The overall picture is that neutrons weakly bound to protons in deuterium are transferred to palladium nuclei, where they are much more strongly bound. In this context beryllium is similar to deuterium, undergoing a photonuclear reaction similar to reaction (2) with an even lower threshold (1.71 MeV), and electrochemical experiments involving beryllium might be instructive.

J. C. JACKSON

Department of Applied Mathematics  
and Theoretical Physics,  
University of Cambridge,  
Cambridge CB3 9EW, UK

1. Fleischmann, M. & Pons, S.J. *electroanal. Chem.* **261**, 301–308 (1989).
2. Jones, S.E. et al. *Nature* **338**, 737–740 (1989).
3. Goldberg, M.D. & Harvey, J.A. *American Institute of Physics Handbook* 3rd edn (McGraw Hill, New York, 1972).
4. Allred, J.C., Armstrong, A.H. & Rosen, I. *Phys. Rev.* **91**, 90–99 (1953).
5. Seagrave, J.D. *Phys. Rev.* **97**, 757–765 (1955).

## Fusion in 1947?

SIR—Fleischmann and Pons<sup>1</sup> claim to have achieved fusion at room temperature in a glass container. I was struck by the similarity to a much earlier experiment by Rayleigh<sup>2</sup>, which demonstrated that a nitrogen (or oxygen or hydrogen) atom impinging on a metal wire in a glass container releases energies ranging up to ~100 eV. At the time, no explanation of the results was advanced, although physicists found no fault with the experiment. In 1957, Burgess and Robb<sup>3</sup> showed that in the presence of traces of oxygen (0.22–0.99 mm Hg) hydrogen atoms will cause the temperature of a metal wire to rise many times above that expected to result from the heat of recombination of H atoms on the wire; this does not, however, explain Rayleigh's results with nitrogen or oxygen. Have they been explained or can we count Rayleigh's experiment as an earlier observation of 'fusion' in a bottle?

KEITH TAYLER

11 Kemp Street,  
Brighton,  
East Sussex BN1 4EF, UK

1. Fleischmann, M. J. & Pons, S. J. *electroanal. Chem.* **261**, 301–308 (1989).
2. Lord Rayleigh *Proc. R. Soc. A* **189**, 296–299 (1947).
3. Burgess, R. H. & Robb, J. C. *Trans. Faraday Soc.* **54**, 1008–1014 (1957).

## Sympatric origins of *R. pomonella*

SIR—Carson raises two important questions concerning the possible recent sympatric origin of the apple infesting form of the maggot fly *Rhagoletis pomonella*<sup>1</sup>. First, he points out that *R. pomonella* may have shifted to domestic apples (*Malus pumila*) from a native host other than hawthorns, such as endemic crabapples. Second, he postulates that different races of *R. pomonella* may exist on different species of hawthorns and that the apple race could have been derived from just one of these 'hawthorn races'. We have addressed these issues in some detail in manuscripts submitted to the journals *Evolution* and *Heredity* and we will therefore highlight only a few of the more crucial reasons why the alternative explanations are unlikely.

The most likely potential native North American hosts for *R. pomonella* other than hawthorns are several species of crabapple (*M. angustifolia*, *M. diversifolia*, *M. glabrata*, *M. ioensis* and *M. coronaria*), none of which supports populations of the fly<sup>2–4</sup>. The reason for this is not clear, but O'Kane<sup>5</sup> has suggested that fruits of native crabapple species are too acidic and ripen too late for *R. pomonella* to complete development before winter. Pree<sup>6</sup> has found a correlation between larval mortality and the total phenol content of fruit from several introduced

species and varieties of crabapples, which suggests that phenolics also contribute to the unsuitability of native crabs as hosts.

Furthermore, in the highlands of Mexico, an isolated population of *R. pomonella* has independently shifted from hawthorns to domestic apples within historical times (apples were introduced into Mexico from Spain in 1522<sup>6</sup>). Both the apple and hawthorn fly populations in Mexico are morphologically distinct from flies in the north-east<sup>1</sup>. Mexican *R. pomonella* could not, therefore, have given rise to the apple race in New England. However, Mexico does not have any endemic species of crabapple<sup>6</sup> (J. Beaman, personal communication). Therefore, in Mexico at least, *R. pomonella* seems to have shifted directly from hawthorns to domestic apples.

It is still possible, of course, that different races of *R. pomonella* exist on different species of hawthorns. Almost 100 different species of *Crataegus*, comprising 19 different species groups, may be endemic to North America<sup>7–9</sup>. But the taxonomic status of most of these *Crataegus* species is questionable as hybridization, polyploidy and apomixis are common in the genus<sup>9–12</sup>. Nevertheless, *R. pomonella* has been reported to attack only a restricted set of *Crataegus* 'species', with infestations confirmed for only 14 endemic hawthorns<sup>1,13–16</sup>. In the north-eastern United States, only hawthorn species that produce relatively large, soft fruits ripening in early autumn are heavily parasitized by the fly<sup>2,4,13</sup>. It therefore seems that *R. pomonella* uses only a few potential hawthorn hosts in the north-east, a trend that is not conducive to the formation of different hawthorn races.

In addition, electrophoretic analysis of *R. pomonella* populations infesting *C. punctata*, *C. brachyacantha*, *C. douglasii*, *C. mollis* and *C. monogyna* from across the United States provides no evidence for genetically differentiated hawthorn races<sup>14–16</sup>. Instead, we find that latitudinal allele-frequency clines exist in the eastern United States among both apple and hawthorn populations (mainly *C. mollis*) for five of the six allozymes showing host associated differentiation<sup>15,16</sup>. The slopes of these allele-frequency clines differ, however, between the two races with the hawthorn race having steeper clines<sup>16</sup>. Inter-host divergence is, therefore, superimposed on north–south clinal patterns of intra-host variation such that the magnitude of genetic divergence between hawthorn and apple flies is a function of latitude. Consequently, different apple populations do not cluster together as a discrete genetic subdivision from populations attacking *C. mollis*, as would be expected if the apple race was formed from a distinct race infesting a hawthorn species other than *C. mollis*.

Instead, the electrophoretic results

suggest that differential selection (affecting development rates for the fly and related to ambient temperature conditions) and differences in host use are key factors responsible for the pattern of allozyme variation for *R. pomonella*. Although further sampling of *R. pomonella* at field sites with sympatric hawthorn species is needed to discount completely the possibility of host-specific hawthorn races, all available data indicate that hawthorn populations represent a single race displaying extensive latitudinal variation.

GUY L. BUSH

Department of Zoology,  
Michigan State University,  
E. Lansing,  
Michigan 48824, USA

JEFFREY L. FEDER\*

Department of Biology,  
Princeton University,  
Princeton,  
New Jersey 08544, USA

STEWART H. BERLOCHER

Department of Entomology,  
University of Illinois,  
Urbana, Illinois 61801, USA

BRUCE A. MCPHERON

Department of Entomology,  
Pennsylvania State University,  
University Park,  
Pennsylvania 16802, USA

D. COURTNEY SMITH

Department of Biology,  
University of Utah,  
Salt Lake City,  
Utah 84112, USA

CHARLEY A. CHILCOTE

Department of Natural Resources,  
University of Michigan,  
Ann Arbor,  
Michigan 48109, USA

1. Carson, H. L. *Nature* **338**, 304 (1989).
2. O'Kane, W. C. *The Apple Maggot* (New Hampshire Exp. Stat. Bull. 171, 1914).
3. Porter, B. A. *The Apple Maggot* (US Dept. Agric. Tech. Bull. 66, 1928).
4. Bush, G. L. *The Taxonomy, Cytology, and Evolution of the Genus Rhagoletis in North America* (Diptera: Tephritidae) (MCZ, Cambridge, Massachusetts, 1966).
5. Pree, D. J. *J. econ. Ent.* **70**, 611–614 (1977).
6. Standley, P. C. *Trees and Shrubs of Mexico* (Smithsonian, Washington, DC, 1922).
7. Fernald, M. L. *Gray's Manual of Botany* 8th edn. (American Book Co., New York, 1950).
8. Correll, D. S. & Johnston, M. C. *Manual of the Vascular Plants of Texas* (Texas Res. Fdn. Renner, 1970).
9. Muniyamma, M. & Phipps, J. B. *Can. J. Bot.* **63**, 1319–1324 (1985).
10. Muniyamma, M. & Phipps, J. B. *Can. J. Bot.* **62**, 2316–2324 (1984).
11. Phipps, J. B. *Ann. Mo. Bot. Gdn* **70**, 667–700 (1983).
12. Phipps, J. B. in *Plant Biosystematics* 417–438 (Academic Press, Toronto, 1984).
13. Wasbauer, M. S. *An Annotated Host Catalog of the Fruit Flies of America North of Mexico* (Diptera: Tephritidae) (Dept. Agric., Sacramento, California, 1972).
14. Berlocher, S. H. thesis, Univ. of Texas at Austin (1976).
15. McPherson, B. A. thesis, Univ. of Illinois at Urbana (1987).
16. Feder, J. L. thesis, Michigan State Univ. (1989).

\* To whom correspondence should be addressed.



# The Right and the wronged

Alex Comfort

**Stalking the Academic Communist: Intellectual Freedom and the Firing of Alex Novikoff.** By David R. Holmes. *University of Vermont/University Press of New England*: 1989. Pp. 288. Hbk \$35, pbk \$14.95. In Britain distributed by Trevor Brown, London, hbk £23.25; pbk £9.95.

ALEX Novikoff was a distinguished American researcher, and a pioneer in cell biology. He was also a Jew who had survived the quota system in force at Columbia University during the 1930s, and, from 1937 to 1945, a Communist. During the McCarthy period he was successively fired from the University of Vermont, widely vilified in the press and, as a long-term sequela, denied a well-earned share in a Nobel prize for his work on cellular organelles. Less resilient or less talented men were destroyed by persecution. Novikoff went on with his work (assisted by Albert Einstein, who personally ensured him a place at the Albert Einstein School of Medicine). He clawed his way back into public esteem, and died a respected and vindicated man, while several of the academic jackals who had attacked him ended with nothing but obloquy.

Books about the anti-Communist crusade of the 1950s make unenjoyable reading. They do, however, have to be read for our protection, exactly as it behoves us to see documentaries about the Holocaust: it could so easily happen all over again, with different labels. David Holmes's scholarly study, which documents the Novikoff story in detail, is particularly valuable if we set it not against the final atrocities of the Nazi regime but against the early days of anti-Semitism and anti-liberalism in Germany, including German academia, which made the gas chambers possible. In the two cases, the disease was identical. Germany succumbed to it; the United States, for reasons creditable to its tradition of individualism, recovered.

In both countries a malign minority was at work, using populist claptrap to inflame opinion and play on fear. In the United States, Tail Gunner Joe McCarthy, the demagogue and fraudster, was simply the front man of an ingenious attack which drew its strength from guilt and fear over nuclear weapons, the anxieties of numerous immigrants, the self-serving of academic and political bigwigs, and the long-nurtured bogey of Communism. Yet the outcome vindicated Lincoln's remark about fooling all the people all of the time. Indeed, the same forces have tried again — are still trying — by different means, but in spite of populist campaigns against 'the L word', and the lobbying of military industry, the United States remains the country of Dan Berrigan and the Christic

Institute as well as the country of Oliver North and Admiral Poindexter. It still, it seems, has what Germany in the 1930s lacked, and blunderbuss anti-Americanism is wide of the mark. This book, indeed, was published for the University of Vermont.

David Holmes does not, in fact, go very deeply into this background, and merely documents exhaustively what happened on the ground to one representative

bolted and his appeals for collegial support encountered a deafening silence. If anyone thinks that this was a feature of a unique period, they should have been at Stanford a few years ago when an equally malicious but non-political attack on a distinguished scientific faculty member — also Jewish — found his colleagues, including his Jewish colleagues, looking at their blotting pads in equal silence.

One likes to think that a British faculty would in both cases have nailed the dean or the president of the university to the faculty club door. But before being smug about it, readers on this side of the Atlantic need to remember that American universities 'respond to market forces', are dependent on scratching for grants and donations, and can accordingly pressure or blacklist even tenured academics if they upset people — precisely the situation, cynics would say, which is being moved

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Senator Joe McCarthy in 1953 and (inset) Alex Novikoff in 1974 on the occasion of being notified of his election to the National Academy of Sciences.

scientist. What he does show, however, is another peculiarly American phenomenon — the craven and rear-protecting attitude of academics faced with attack from outside. The United States has a long tradition of distrust for eggheads — a Martian anthropologist landing in Peoria might opine that technology is lionized but learning is despised, and most American academics I have met are aware that public toleration is fragile. The prejudice is an old one, part of the mask of hard-headedness. Since the days of the Novikoff case it has been changed and, if anything, reinforced by the 'mad scientist' fallout from nuclear weapons and the like.

McCarthy caught this quietist and rather insecure academic community unprepared and apprehensive. Attacked in this way it had no stomach for a concerted counter, and its unwillingness to take public positions compounded the panic. While political climbers and operators like Borgmann, the social scientist dean of faculty who aimed to make the University of Vermont "an agency of state", encouraged the witch hunt, spouting sweet reasonableness, Novikoff's colleagues incontinently

surreptitiously into position in Britain.

All these deeper issues apart, Holmes has provided a fascinating chronicle of malice, hypocrisy and, of course, courage. If Novikoff was naive about Stalinist Communism — and in 1937 he was by no means alone — he more than made up for it by his Job-like persistence in the face of gross injustice. By 1957 he had been investigated nine times by various security organizations. Even Senators disgusted with Tail Gunner Joe and working for his downfall would not speak up publicly for Novikoff. In fact, it was really only after his death that the condolences and appreciations become fulsome. (Not, it appears, that Novikoff was an easy man to get on with. He was seen by intimates as overbearing and on occasions rash, but this was perhaps as well for him in that it provided the iron which he needed to see things through.) Holmes's book becomes, by reason of its careful detail, a blacklist in itself — of self-proclaimed liberals who did nothing about the witch hunt out of concern for their own skins.

There is, as always, another side to this and similar unedifying passages of history — exactly as there are factors in German

Popperfoto

history which provided a handle for Hitler. Fear of the Left, which threatened personal wealth, had long been a leading player at the American table. When Novikoff joined it, Hitler had a vocal and influential fan club both in Britain and the United States. Suddenly the game was transformed and the Soviets became allies. The Bomb then restored the ability of the American Right to sleep soundly at night, but — hey presto! — the savage Communists had one too, and domestic traitors had helped them get it. Fear created as a political instrument turned to paranoia, and it took some years for the Constitution to regain its authority.

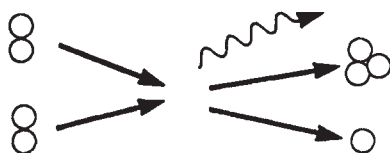
It is this background in American self-perception that lies behind the more parochial events documented by Holmes. The historian needs both perspectives — but it is salutary and right to see the outcome of history at the personal level, in the chronicle of its effects on one man. This Holmes enables us to do: which is why his book is appointed to be read not only in the United States but in British universities now under threat of 'reform'. □

*Alex Comfort, The Windmill House, The Hill, Cranbrook, Kent TN7 3AH, UK, is a scientist, poet and novelist. Best known of his books is The Joy of Sex; less well known is that he refused military service in the Second World War.*

Pd



## Palladium Wire



**Small Quantities —  
Low Price**  
*Call Today For Quote*

**Marshall Laboratories**  
5854 Rawhide Ct.,  
Boulder, CO 80302 USA  
Reader Service No.7

Telephone:  
(303) 442-9004 • (303) 442-0156  
FAX (303) 440-3588

## Beginnings in an Eastern Eden

John D. Barrow

**Creation of the Universe.** By Fang Li Zhi and Li Shu Xian. *World Scientific*: 1989. Pp.180. Hbk \$41, £26; pbk \$25, £17.

COSMOLOGISTS have a nagging history of conflict with authority. If your name is Aristarchus, Bruno, Galileo, Copernicus or Sakharov, then you know what I mean. The principal author of this book, Fang Li Zhi, has a similar claim to fame in the West. Until 1987 he was vice-president of the University of Science and Technology of China, but early in that year he was removed from his post and banished to the Beijing Astronomical Observatory after having been accused of stirring up student demands for greater freedom and democracy in China. More recently he was prevented from meeting President George Bush of the United States during his visit to China.

The authors remark in their preface that only by surmounting considerable hurdles has it been possible that their book could be published at all. Yet current events in China may well bring concepts such as 'freedom and democracy' back into fashion and return Fang Li Zhi to unsuspected prominence. In the meantime, he has produced a number of popular accounts of modern cosmology. *Creation of the Universe*, written with Li Shu Xian, is the most recent of them and offers an account suitable for introductory reading by science students or general readers with some background in mathematics and physics. The mathematics is elementary and is frequently alleviated by the liberal use of good diagrams and elegantly drawn cartoons; the English is for the most part good enough not to distract from what the authors are trying to convey.

The content is a sandwich in which the substantial central section of six chapters contains clear accounts of the dark matter problem, thermodynamics, the hot phase of the early Universe, the origin of the elements, matter and anti-matter in the Universe, and inflationary expansion. Occasionally, there are minor lapses — magnetic monopoles are produced in the inflationary Universe, but they are swept beyond our visible horizon; the history of the prediction of microwave background by Gamow is imaginary; the resolution of Olbers' Paradox is simplistic; the strong conclusions drawn about elemental abundances ignore the possible contributions by a pregalactic population of very massive stars; and the discussion of the geometry of the Universe, which argues

that there are only two possible cases, ignores the simple euclidean possibility altogether.

Nevertheless, there are also many commendable features. There are no unnecessary complexities or mathematical smokescreens, and up-to-date accounts of the relevant observational data are given. Whenever ideas from particle physics are required, as for example in the discussion of the origin of baryon asymmetry in the early Universe, they are economically introduced.

Many of these explanations and simplified derivations will be familiar to the intended audience, so it is a pity that the authors provide no references to other works. Given that their aim could have been to do nothing more than initiate the student into the basic physical ideas of modern cosmology, it seems perverse not to have directed the reader towards more substantial sources.

The core of the book is sandwiched between two rather wobbly sections which cannot really be recommended. Both are far too ambitious and are unsuccessful, although for quite different reasons. In the erratic introductory chapter, the authors attempt to explain the whole sweep of structures in the Universe as a generalization of Laplace's picture of the origin of the Solar System in which the flattened disk of the solar nebula arises because of rotation. The subsequent claim that spheroidal galaxies and galaxy clusters thus have non-spherical shapes because of rotation is not correct.

Also distinctly dodgy is the ensuing argument that small-scale structures in the Universe have symmetrical forms because they have completed gravitational collapse, whilst ragged superclusters of galaxies are asymmetrical because they have not. In fact the very opposite is more likely to be true. In the absence of pressure, gravitational collapse magnifies deviations from spherical symmetry; large-scale structures are more likely to remain spherical if they are still being dragged along by the expansion of the Universe.

The final sections of the book contain a generally good treatment of various 'anthropic' consequences of physical constants. It is, however, spoilt by several pages of fairly vacuous 'philosophical' discussion of 'being' and 'non-being' and other such matters that only sound as if they add to our knowledge about the creation of the Universe.

Despite these occasional problems, the credits end up outweighing the debits. Here is a worthwhile elementary treatment of the cosmology of the early Universe written with a liveliness and simplicity that will surely encourage students to pursue the subject further. □

John D. Barrow is in the Astronomy Centre, University of Sussex, Brighton BN1 9QH, UK.



# English-style geometries

Ivor Grattan-Guinness

**Mathematical Visions: The Pursuit of Geometry in Victorian England.** By Joan L. Richards. *Academic*. 1988. Pp.266. \$34.95, £22.

IN THIS revised version of her doctoral dissertation defended at Harvard University, Joan Richards covers an important and rather individual stream of work in English Victorian mathematics: the concern with geometries of various kinds (euclidean, non-euclidean and projective), and the relations between them. Starting out from the renaissance of mathematics at Cambridge in the 1810s, she concentrates on the second half of the century, with chapters devoted to the philosophy of geometry, the popularity of its non-euclidean variety, the role of the projective version, the disputes over teaching the euclidean type in schools, and the early researches of Bertrand Russell.

Russell was not the only person to apply logic to geometry, for the matter involved various issues of consistency and (in)completeness of theories. Richards's own use of vocabulary is less clear than it might be: especially her use of 'analysis' (which is sometimes the mathematical subject and otherwise the proof-method); and of 'validity' of theorems, which can either be truth-values within a theory or their verisimilitude as descriptions of space or of empirical phenomena.

The book shows a high degree of scholarship, with a broad range of sources used, including Parliamentary papers and *Nature* itself (whose origin is falsely attributed on p.164 to T.H. Huxley). The general context of scientific thought is much better outlined than is normally the case in books on the history of mathematics, although I find her characterization of the science of the period as "post-Darwinist" a little narrow, because the life and Earth sciences *in general* provided the mathematicians' stimuli for comparison.

Standard of production is also good. Footnotes are literally so, and the references are repeated in the bibliography at the end. A few items are significant by their absence, however, and when taken into account they somewhat temper the portrait that is presented.

For example, the disputes over school education were richer in content and context than is suggested in Chapter 4. And as Richards hints on pp. 179–180, both sides of the dispute were ignorant of studies of euclidean geometry being pursued on the Continent, which exposed its logical weakness. By 1902 Russell

could point out some examples of the modern critical position, in a paper entitled "The Teaching of Euclid".

At the higher level of instruction, especially at Cambridge, the philosophical issues hinged around the principle that 'training the mind', as conceived in the 1820s, was designed to do no more than that, rather than invite the really enquiring mind to pursue original research; and yet important scientists did emerge out of this system. Both sides of the paradox have been exhibited elsewhere, but the point itself is not quite made in this book.

In starting with the Cambridge reform, Richards follows the usual historical interpretation of changes in the British Isles. However, in a volume currently in production with Cambridge University Press, N. Guicciardini has shown that various other institutions in Britain had initiated reforms before that. In England these were the military schools; and thus one is led to wonder as to their roles in nineteenth-century English geometry. The question is not raised, although it is clear that there

was considerable interest at least in descriptive geometry.

Finally, the (implicit) characterization of Victorian mathematics as preoccupied with geometry ignores another face, partly in contradiction with that shown here. *Mathematical Universes: The Pursuit of Algebra in Victorian England* could be the title of a book which covered as much mathematical ground and involved as many notable figures as does this one: general algebras, differential operators, functional equations, linear algebra and probability theory were especially favoured topics. Richards discusses aspects of the first of these areas in places, mainly in the context of A. de Morgan, of whom she is currently writing a biography. One hopes that next she will tackle the Victorian interest in algebras more deeply, for few historians possess her knowledge of that period. □

Ivor Grattan-Guinness, *Middlesex Polytechnic, Enfield, Middlesex EN3 4SF, UK*, is the Immediate Past-President of The British Society for the History of Mathematics.

## Little wonders

Tom Fenchel

**Microbial Ecosystems of Antarctica.** By Warwick F. Vincent. *Cambridge University Press*. 1989. Pp.304. £37.50, \$75.

MOST people associate Antarctic life with the large populations of penguins, seals and whales that are sustained by immense numbers of krill and other zooplankton. Yet the harsh environment of most of the continent allows only for microbial life; in the absence of grazing animals, microbes often reach high enough population densities to become macroscopically visible as brown-stained ice, red-coloured snow fields and black crusts on soils. As documented by the 500 or so references in this book, over the past two decades microbiologists have been busy at work on the wide variety of scientific questions involved.

In the core of the book, Warwick Vincent describes the different habitats: snow and ice, sea ice, the open ocean, lakes, streams, soils and rock surfaces. For each of them, environmental factors, the composition of the microbial communities, and their importance for the ecosystems and biogeochemical activities are discussed. We are presented with new insights into Antarctic marine planktonic food chains; for example, it has been found that, during the summer, the receding sea ice creates vertically stabilized and nutrient-rich water masses, and is thus responsible for the bulk of the phytoplankton production which sustains the large populations of marine invertebrates, birds and mammals. The descrip-

tions of microbial life in the most extreme environments — snow, ice bubbles, hyperhaline lakes (which remain unfrozen at  $-30^{\circ}\text{C}$ ) and exposed rock surfaces — are fascinating. It is, though, a pity that a section comparing Antarctic microbial communities with their Arctic counterparts was not included.

The other main section of the book treats adaptations to the environment at the cellular level. Survival and metabolic activity at freezing temperatures is an obvious problem, but Antarctic microbes may also be exposed to dehydration and to extremely high salinities (the combined effect of freezing and evaporation). A final chapter is devoted to the effects of human activities, for example pollution and the introduction of xenobiotic microbes, and to the possible practical use of Antarctic microbes. Also included are a glossary, and two appendices on Antarctic climates and tables of environmental data.

This is a well-written, well-illustrated and authoritative volume. As well as biologists concerned with the Arctic or Antarctic, anyone interested in bacteria and protists in nature will find it well worth reading. □

Tom Fenchel is in the Marine Biological Laboratory, University of Copenhagen, Strandpromenaden 5, DK-3000 Helsingør, Denmark.

### New in paperback

- *The Connection Machine* by W. Daniel Hillis. Publisher is MIT Press, price is \$13.95, £12.50. For review see *Nature* 321, 126 (1986).
- *Klaus Fuchs, Atom Spy* by Robert Chadwell Williams. Publisher is Harvard University Press, price is £9.95, \$12.95.
- *Artificial Intelligence: The Very Idea* by John Haugeland. Publisher is MIT Press, price is \$9.95, £8.95. For review see *Nature* 319, 185 (1986).

# Law breaking

J. H. Mulvey

**The New Physics.** Edited by Paul Davies. Cambridge University Press: 1989. Pp. 516. £30, \$49.50.

CLAIMING an omniscience sometimes irritating to other scientists, physicists regard the whole of the physical world as their domain. They dare to explain the birth of the Universe, expect to find simplicity in the midst of evident complexity, and dissect matter to find its smallest constituents and to uncover the laws ruling its behaviour. The discovery of the electron, the atomic nucleus and the laws of quantum mechanics answered the questions posed by Mendelev's periodic table of the elements and provided the basis for our understanding of almost all modern science and technology. The first 30 years of this century, which also saw the advent of Einstein's special and general theories of relativity, are still looked upon as the golden age of physics.

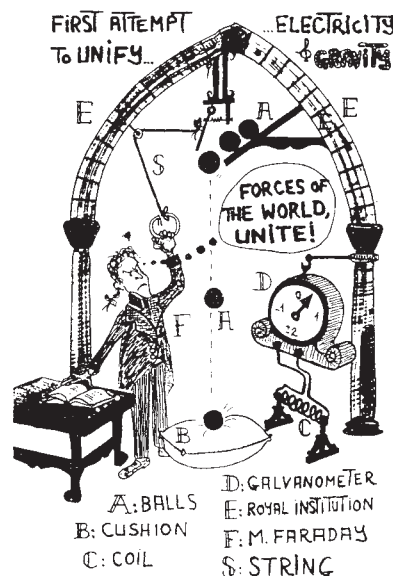
Paul Davies's objective in compiling *The New Physics* is to advertise the advances of more recent decades, taking us on to new frontiers. To span an ambitiously broad sweep of topics he has the help of 18 contributors, all of whom are distinguished in their fields and who write in a welcome diversity of styles: flamboyant, systematic, excited, thoughtful, fanciful. The chapters are complete in themselves leading inevitably to some overlap, but this is often helpful as some authors are more successful than others in explaining unfamiliar terms and concepts. The level is generally that of *Scientific American*, though some contributions have tougher sections and several use mathematical language.

The first section of the book deals with physics on the scale of the Universe, where gravity reigns. The last concentrates on particle physics and the tantalizingly close approaches to a unified theory of the forces of nature. In the middle we find physics on a more human scale, dealing with systems amenable to control and manipulation in the laboratory but often showing great complexity of behaviour.

Clifford Will, Chris Isham, and Alan Guth and Paul Steinhardt review general relativity, the still elusive quantum theory of gravitation, and the inflationary version of the Big Bang model of the early Universe. The latter provides a mechanism for matter and energy to emerge out of nothing — "perhaps the ultimate free lunch". Stephen Hawking's short essay ponders the "edge of spacetime", contemplating the curvature of time as well as of space so that to ask "What came before the Big Bang?" would be like asking to go one mile north of the North Pole. In a

well-illustrated 'mini-monograph' Malcolm Longair describes and interprets the vast range of observational data that has transformed our view of the Universe.

The title, "Critical Point Phenomena: Universal Physics at Large Length Scales" does not immediately evoke the flavour of fashionable 'new physics', but Alastair Bruce and David Wallace pick up one of the unifying threads running through the book. They demonstrate the significance of relationships between the lengths, or scales, characterizing a physical system and introduce the ideas underlying the



Michael Faraday at the Royal Institution, a cartoon by A. de Rujula (reproduced from *The New Physics*).

mathematical technique known as the renormalization group.

David Thouless describes the mobilization of armies of atoms or electrons forming linear or two-dimensional arrays in physicist-engineered materials that have many novel properties. The world of low-temperature physics, at temperatures close to absolute zero where quantum mechanics rules on a macroscopic scale, is the topic for Anthony Leggett. More examples of elegant physics are described by Peter Knight writing on quantum optics; they range from the stimulation of great orchestras of atoms to emit intense beams of laser light to the observation of a lone atom 'winking'.

The development of methods for the treatment of nonlinear processes forms a distinct branch of the new physics. Gregoire Nicolis discusses far-from-equilibrium systems and self organization, as in the regular patterns of convection in a heated liquid. Such complex systems can evolve into random, unpredictable motion and this is the subject for Joseph Ford, a self-proclaimed "disciple of chaos".

Abner Shimony, writing more in the style of a philosopher than a physicist, reviews the conceptual foundations of quantum mechanics. Frank Close explains

the basis for our belief in the unobserved, perhaps unobservable quarks which are, with the electron, today's ultimate building blocks of matter. Howard Georgi makes two lively contributions. One is on Grand Unified Theories (GUTs); these attempt to find a theoretical framework unifying quantum chromodynamics (QCD), the theory describing the force between quarks, with the electroweak theory which unites electromagnetism and the weak interaction (responsible, *inter alia*, for radioactivity). In his second piece Georgi describes his views on the status of quantum field theories. Here he warns theoreticians against extending their speculations into energy regions too far beyond the discipline of experiment in a search for mathematical elegance.

Three of the four basic forces of nature are now understood and described by quantum field theories. They are all 'gauge forces', a term explained by John Taylor who goes on to review their quite different characters: the infinite range of electromagnetism, the short range of the weak force, and the extraordinary feature of quark confinement in QCD.

The last word is given to Abdus Salam. Fancy as ever unconstrained, he contemplates 'theories of everything' based on the concept of supersymmetric strings in as many as 26 dimensions, and speculates on the feasibility of lunar-encircling particle accelerators. After describing some of the theoretical results, he makes a remark which can stand for the book as a whole: "One can only hope nature is aware of our work on all this".

Two impressions stand out: the remarkable unity of the new physics, and the role of curiosity as the vital driving force of discovery. Shared concepts and methods link the different sectors in ways scarcely imagined 20 years ago; examples are the inflationary model of the early Universe based on ideas from GUTs, and the significance of dimensional scales and the renormalization process which underlie approaches to problems as varied as symmetry breaking in gauge theories, critical point phenomena, self-organization, and the development of chaos. Applications are hardly, if ever, mentioned as a motive, but the 'useful products' already realized are of immense value and more will come. Here are two lessons for those, even to be found among scientists, who would manage and direct science towards 'purposeful' ends.

This is an attractively presented book, but it is one to be read and thought about, rather than merely looked at. Non-scientists will get a good sense of some of the new ideas, and scientists — physicists included — can learn how very far physics has extended its dominion □

J.H. Mulvey is a Senior Research Officer in the Nuclear Physics Laboratory, University of Oxford, Keble Road, Oxford OX1 3RH, UK.



# Meteoritic silicon carbide and its stellar sources; implications for galactic chemical evolution

Tang Ming<sup>\*‡</sup>, Edward Anders<sup>\*</sup>, Peter Hoppe<sup>†‡</sup> & Ernst Zinner<sup>†</sup>

<sup>\*</sup> Enrico Fermi Institute and Department of Chemistry, University of Chicago, Chicago, Illinois 60637-1433, USA

<sup>†</sup> McDonnell Center for the Space Sciences and Department of Physics, Washington University, St Louis, Missouri 63130, USA

Interstellar silicon carbide grains in meteorites provide a novel means for studying the carbon-star population of about  $5 \times 10^9$  years ago. Their  $^{12}\text{C}/^{13}\text{C}$  ratios differ greatly from the solar value but resemble those of present-day carbon stars, implying little change in the galactic  $^{13}\text{C}$  inventory. Isotope data on nitrogen and silicon suggest that the silicon carbide grains come mainly from red giants, with small contributions from novae.

CARBONACEOUS chondrites contain several p.p.m. of SiC (refs 1–3), of highly anomalous isotope composition<sup>4</sup> and hence of interstellar origin. These grains condensed in stellar outflows and were trapped in meteorites when the Solar System formed. We report data on new SiC samples, including single crystals 3 to 12  $\mu\text{m}$  in size—much larger than predicted by current models of circumstellar grains—that were large enough for individual analysis by ion microprobe. Overall, the isotope ratios of N, C and Si vary by factors of 100, 16 and 1.3, respectively. The Si ratios suggest that the SiC comes from at least four separate stars, and the C and N ratios indicate that carbon-rich giants are the main source, with lesser contributions from novae and perhaps Wolf-Rayet stars. Interestingly,  $^{12}\text{C}/^{13}\text{C}$  ratios of SiC resemble those of present-day carbon stars, despite the age difference of  $\sim 4.6 \times 10^9$  yr. SiC in meteorites is only  $\sim 10^{-4}$  times as abundant as predicted for fresh interstellar matter, implying destruction or an initial deficiency in protosolar matter.

Three SiC fractions were isolated from the Murchison C2 chondrite<sup>5,6</sup>. Silicates were dissolved by HF–HCl, and carbonaceous matter by  $\text{HClO}_4$ , which left a residue of spinel, diamond and SiC. After removing diamond as a colloid in 3 M  $\text{NH}_3$ , the remaining material was divided by sedimentation into four grain-size fractions. The fractions 0.03–0.2  $\mu\text{m}$ , 0.2–2  $\mu\text{m}$  and 2–10  $\mu\text{m}$  were heated with  $\text{H}_3\text{PO}_4$  to dissolve spinel, which gave SiC fractions HM, HN and HO (3.3, 3.6 and  $\leq 0.3$  p.p.m.; their SiC contents were 76, 94 and  $\sim 10$ –50%). Noble-gas and complete ion-probe analyses will be reported by Zinner *et al.* (in preparation).

Ion-probe isotope analyses of C, N and Si were done with a  $\text{Cs}^+$  beam of 3–5- $\mu\text{m}$  diameter, generally on aggregates comprising  $\sim 10^1$ – $10^3$  grains, but for HO also on 12 individual grains, some of which are shown in Fig. 1. Techniques were similar to those of ref. 4, and will be described in detail by Zinner *et al.* (in preparation).

## Number of stellar sources

The Si data for HO (Fig. 2) are highly anomalous, rarely approaching terrestrial composition (at the origin) and usually lying above the terrestrial mass fractionation line (which is

shown only for positional reference). Sample HN shows a similar distribution, whereas the points for fine-grained HM cluster more tightly<sup>6</sup>. An important question is the number of independent isotope components (endmembers) present. On such a three-isotope plot with a common denominator, mixtures of two components lie on a straight line joining them both, whereas mixtures of  $n$  component lie within an  $n$ -sided polygon of minimum area whose perimeter passes through or encloses the  $n$  components.

A trivial solution exists in the form of a triangle large enough to enclose all points, corresponding to  $\geq 3$  components at or beyond the vertices<sup>4</sup>. But a more meaningful answer comes from the single grains, which, being single crystals (Fig. 1 and transmission electron microscopy data), must be pure components rather than mixtures, and thus represent the composition of their stellar formation sites. For a minimum estimate, we pool those whose error bars overlap within  $2\sigma$ . This leaves five discrete components, plus at least one other represented by the aggregate that lies outside the polygon defined by these five components (Fig. 1 of ref. 4).

Silicon is unaffected by H- or He-burning under non-explosive conditions, and if none of the SiC comes from exploding stars (novae or supernovae), then these  $\geq 6$  Si components would represent  $\geq 6$  discrete stars. However, as we shall see from the C and N data, some of the SiC grains do show a nova signature, and so the minimum number of parent stars is smaller.

Three grains (Group 2, at 0, 0 in Fig. 2) have essentially terrestrial Si composition, but none are contaminants: all have grossly anomalous C and/or N. Curiously, these three grains have very fresh surfaces, either crystal faces (Fig. 1, upper left) or conchoidal fracture surfaces (Fig. 1, upper right). The other two grains in Fig. 1 (lower left and right) have fuzzy, 'weathered' surfaces, and their Si isotope compositions are anomalous

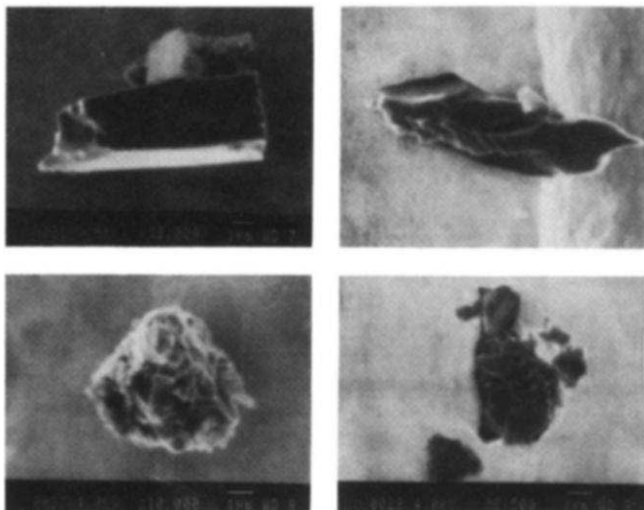


FIG. 1 Large SiC grains from the Murchison C2 chondrite. Grains at the bottom have a 'weathered' appearance.

<sup>‡</sup> Present addresses: Department of Management Sciences, University of Minnesota, Minneapolis, Minnesota 55455-0424, USA (T.M.); Physikalisches Institut der Universität, CH-3012 Bern, Switzerland (P.H.).

(Groups 3 and 1). It remains to be seen whether this apparent correlation between freshness and Si isotope composition is borne out by future work. Clayton<sup>7</sup> has predicted that younger grains should be enriched in the secondary isotopes  $^{29}\text{Si}$  and  $^{30}\text{Si}$ , which formed later, on average, than the primary isotope  $^{28}\text{Si}$ . If weathering is a measure of age, then these five grains fail to show this trend, with one weathered grain (Group 3) enriched rather than depleted in both heavy isotopes, and the other (Group 1) depleted only in  $^{29}\text{Si}$ .

Interestingly, the Si in average meteoritic SiC is heavier than terrestrial or meteoritic (silicate) Si, having  $\delta^{29}\text{Si} \approx +26$  and  $\delta^{30}\text{Si} = +35\%$ . This may reflect differences either in stellar source (carbon-rich stars for SiC, oxygen-rich stars for silicates) or in age: in terms of Clayton's model, the SiC would be younger than the silicate grains that supplied most of the Si to the Solar System. Indeed, the mean cosmic-ray exposure age of meteoritic SiC is only  $41 \pm 20$  Myr (ref. 5), much shorter than the predicted mean lifetime of refractory interstellar grains,  $\sim 500$  Myr (ref. 8).

### Comparison with carbon stars

As SiC apparently comes mainly from red giants, we compare the  $^{12}\text{C}/^{13}\text{C}$  ratios of SiC with those of red giants (Fig. 3). The top panel shows 32 M, MS, S and Ba stars whose low  $^{12}\text{C}/^{13}\text{C}$  ratios reflect mainly H-burning by the CNO cycle, whereas the second panel shows 38 carbon stars (classes SC and N), which have produced additional  $^{12}\text{C}$  by He-burning<sup>9,10</sup>, thus elevating  $^{12}\text{C}/^{13}\text{C}$  and raising C/O above the threshold value ( $\sim 1$ ) for formation<sup>11</sup> of SiC. The next three panels show the data for SiC of decreasing grain size, including the Murray CF and CJ samples from Zinner *et al.*<sup>4</sup>.

The distribution for the coarsest-grained sample, HO, is rather similar to that for carbon stars, peaking between 50 and 60 and covering roughly the same range. The highest  $^{12}\text{C}/^{13}\text{C}$  value of 159 exceeds the (known) carbon-star range as well as the highest value of 120 previously found in meteorites<sup>12</sup>, and presumably

reflects a very large contribution from He-burning sources. As this analysis was for an aggregate rather than a single grain, some individual grains must have higher values still.

The range for the remaining samples decreases with grain size, because of the larger number of grains per analysis. Of the finest two samples, CF has a distinctly higher mean than HM (52 as opposed to 39), apparently reflecting contamination by microdiamonds<sup>4,13</sup> with  $^{12}\text{C}/^{13}\text{C} = 93$ . Stepped combustion<sup>13</sup> shows that the main SiC component in CF actually has  $^{12}\text{C}/^{13}\text{C} = 37.6$ , close to the value for HM.

The (mass-weighted) mean  $^{12}\text{C}/^{13}\text{C}$  ratios for SiC are  $44 \pm 3$  for all five samples or  $41 \pm 3$  for the Murchison samples only. Both are remarkably close to the mean for the 38 carbon stars,  $(45 \pm 4)$  and for the present-day interstellar gas<sup>14</sup>  $(43 \pm 4)$ , but are decidedly lower than the Solar-System ratio of 89. The difference between the last two is often attributed to galactic evolution over the past  $4.6 \times 10^9$  yr, which produced  $^{13}\text{C}$  preferentially (compare Table 13 in ref. 14). But if that explanation is correct, then why does  $4.6 \times 10^9$ -yr-old SiC ( $44 \pm 3$ ) match present-day carbon stars ( $45 \pm 4$ ) and interstellar gas ( $43 \pm 4$ )?

One possible reason is that the three data sets are not truly comparable, and thus agree only by chance. Meteoritic SiC is a limited sample from one corner of our Galaxy, the carbon stars in Fig. 3 are a selected sample of two spectral classes that cover only part of the range of SiC-producing stars and are observed at different stages of carbon production, and the interstellar gas is a more comprehensive mixture, coming mainly from stars whose C/O is too low for condensation of solid C or SiC. Another possible reason is that the interstellar concentration of nucleosynthesis products apparently has not changed systematically in the last few billion years, judging from the mean Fe/H ratio of young globular clusters (compare Fig. 3 of ref. 15). A further possibility is that the Sun is anomalous, having been enriched somehow<sup>16</sup> in  $^{12}\text{C}$ ,  $^{16}\text{O}$  and  $^{20}\text{Ne}$ . What is needed for the present comparison is an average integrated over stellar

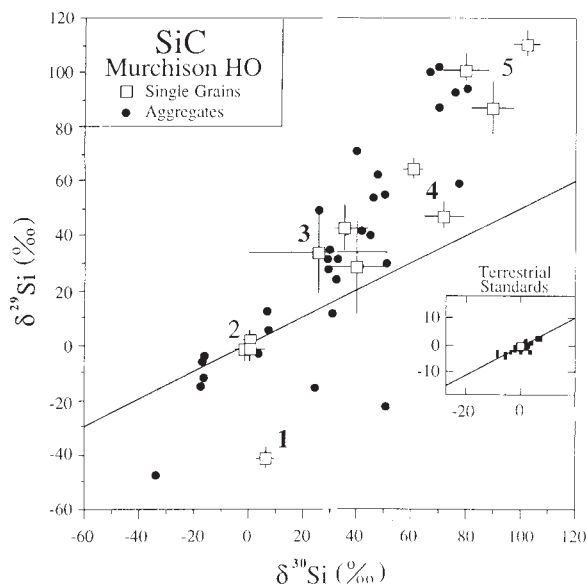


FIG. 2 Isotope composition of Si, relative to terrestrial standard.  $\delta^i\text{Si} = 1,000 (R_{\text{sample}}/R_{\text{standard}} - 1)$ , where  $R = {}^i\text{Si}/^{28}\text{Si}$ . Single grains are pure components but with some surface debris, whereas aggregates are mixtures. The 12 single grains fall into 5 groups that presumably represent  $\geq 5$  discrete stars. Error bars are shown only for single grains; they are dominated by counting statistics, reflecting the amount of material analysed. In addition there is an instrumental mass fractionation, as indicated by the spread of data for the terrestrial SiC standard (inset). Error bars for aggregates are comparable but in contrast to single grains, they also include true compositional variation in addition to analytical precision.

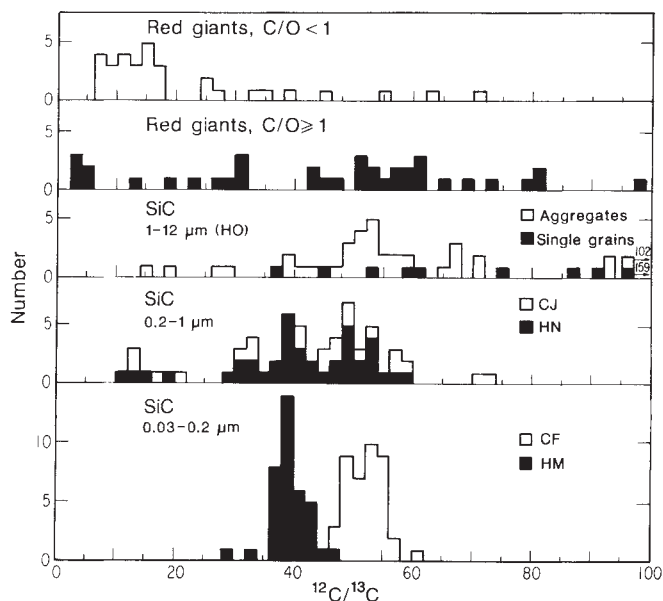


FIG. 3  $^{12}\text{C}/^{13}\text{C}$  ratios of red giants and meteoritic SiC. The SiC data include previous measurements<sup>4</sup> on two Murray separates, CF and CJ. The finer-grained meteoritic samples cluster around the mean because of averaging, but the coarsest fraction, HO, has a distribution very similar to that of carbon stars ( $\text{C/O} \geq 1$ ), which are known to condense SiC in their outflows. This similarity is surprising in view of the  $4.6 \times 10^9$ -yr difference in age, because preferential production of  $^{13}\text{C}$  in the evolving Galaxy should cause the  $^{12}\text{C}/^{13}\text{C}$  ratio to decline with time. For stars, only high-precision data were used, based on Fourier-transform spectroscopy of infrared lines<sup>9,32,37</sup>.



mass and over  $^{12}\text{C}/^{13}\text{C}$ , C/O and production of SiC in time. An exact solution may be decades away, but even a first approximation may provide some new constraints for models of galactic evolution.

### Novae and other sources

We can look for contributions by other processes and other stars, using a plot of nitrogen against carbon isotope ratios (Fig. 4). The field is divided into quadrants by the solar ratios, and the dominant nucleosynthetic processes are indicated for each quadrant. This diagram shows qualitatively which processes are needed to change one isotope composition to another, but as the initial composition was not necessarily solar, the dividing lines between quadrants are not as sharp as implied by the solar ratios.

Most of the grains lie in the upper left quadrant to the right of the equilibrium value for H-burning by the CNO cycle (shown by a cross). As discussed earlier, this rightward shift reflects He-burning and dredge-up of  $^{12}\text{C}$ . The variation in  $^{14}\text{N}/^{15}\text{N}$  is large even for the fine-grained aggregates from HM, suggesting the presence of quite extreme compositions. The single grains show no correlation between Si-group (indicated by numerals) and C, N: two grains with virtually identical C, N belong to different Si groups (3 and 5), whereas the three group 2 grains are widely separated, as are the two group 5 grains. An obvious reason is that the C and N compositions in a red-giant atmosphere change continually by repeated dredge-up of material from the He-burning shell; thus successive crops of SiC will have constant Si but variable C and N.

The points in the lower left quadrant are enriched in  $^{13}\text{C}$  and  $^{15}\text{N}$ . The latter is made by very few stellar processes, mainly by explosive H-burning in novae, which also produces  $^{13}\text{C}$ . Interestingly, novae have been invoked<sup>17</sup> as the source of the Ne-E(H) component<sup>4-6</sup> in SiC, that is, a nearly monoisotopic  $^{22}\text{Ne}$  that presumably comes from decay of  $^{22}\text{Na}$  (half-life of 2.6 yr). Because of its short half-life, the  $^{22}\text{Na}$  must have been produced and trapped very quickly, that is, under explosive conditions<sup>17,18</sup>. Novae also have C/O ratios that are high enough for condensation of SiC.

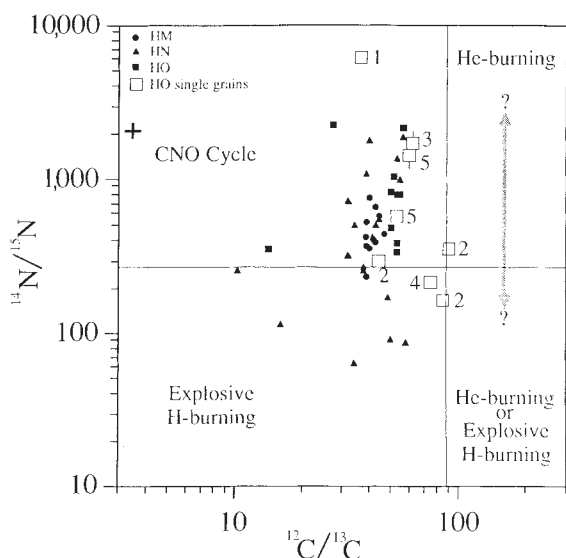


FIG. 4 C, N isotope ratios of meteoritic SiC, compared with solar ratios (thin lines). Numbers refer to Si groups from Fig. 2. Most SiC points lie in the upper left quadrant, where H-burning by the CNO cycle and He-burning are the principal processes. Several points lie in the lower left quadrant, where the high abundance of  $^{15}\text{N}$  suggests a contribution from explosive H-burning, presumably in novae. Double-pointed arrow represents an aggregate of unknown nitrogen isotope composition.

The lower right and upper right quadrants are almost unpopulated, with only one aggregate (denoted by a double arrow), of unknown N composition, falling well within this region and one or two others marginally so. The  $^{12}\text{C}/^{13}\text{C}$  ratios lie near or beyond the highest value seen in carbon stars (98), but in principle there is no reason why such ratios cannot be attained at advanced stages of He-shell burning<sup>9,10,19</sup>. Two other potential sources, both with C/O > 1, are the carbon shells of Type II supernovae and Wolf-Rayet stars of classes WC 8-10 (ref. 20). Under certain conditions,  $^{15}\text{N}$  can also be produced in explosive H-burning<sup>21</sup>, but it is not yet clear what conditions and what stellar objects will simultaneously account for the  $^{12}\text{C}/^{13}\text{C}$ ,  $^{14}\text{N}/^{15}\text{N}$  and C/O ratios. A clearer picture may emerge when more extreme grains are found in these quadrants, especially if their noble-gas components are also characterized. But it is already obvious that the SiC is a melange from various sources.

Explosive H-burning also changes the isotope composition<sup>22</sup> of Si, and thus only the Si components associated with C and N compositions in the upper left and upper right quadrants (where hydrostatic rather than explosive burning dominates) can be safely regarded as coming from single stars. Of the single SiC grains, one falls into a quadrant in which explosive burning dominates, whereas one other cannot be located on Fig. 4 for lack of N data. That leaves at least 4 and possibly 6 or more individual stars.

The 30% dispersion of the Si components (Fig. 2 and ref. 4) should be a good measure of the isotopic heterogeneity of the interstellar medium, on the scale of at least individual stars. This is a new and important constraint for theories of the interstellar medium. A second constraint is the very large maximum size of the SiC (Fig. 1), exceeding previous estimates by one to two orders of magnitude. Perhaps conditions necessary for the formation of circumstellar grains have been too pessimistically estimated.

### Rarity of SiC in the Solar System

The extreme rarity of SiC in meteorites is a major surprise. Carbon stars are thought to contribute about 1/3 of the total mass ejected from stars<sup>20</sup>, and because all of their Si is expected to condense as SiC (ref. 11), it is surprising that even primitive (C2) chondrites contain only  $4 \times 10^{-5}$  of their Si as SiC. (Less primitive meteorites would have lost part or all of their SiC by thermal metamorphism, if they ever contained it.) Three destruction processes are obvious candidates.

**Supernova shocks.** Refractory interstellar grains usually are destroyed by supernova shocks on a timescale of  $\sim 500$  Myr, (ref. 8), but because there are indications for a late supernova contribution to the early Solar System<sup>16,24,25</sup>, the SiC grains in the protosolar neighbourhood may have been largely destroyed by such a supernova. However, the arguments for a nearby supernova are not compelling<sup>26</sup>, and destruction of SiC by the required factor of  $< 10^{-4}$  is hard to reconcile with the high abundance of microdiamonds<sup>27</sup> ( $\sim 10^{-3}$  of the cosmic complement of C) that bear a supernova signature in the form of Xe-HL.

**Selective interstellar process.** SiC has been observed only in circumstellar, not interstellar, dust. This deficiency may be observational rather than real, but if accepted at face value it implies a rapid, selective destruction process, perhaps reflecting the chemical instability of SiC in a medium of low C/O. However, depletion by a factor of  $\sim 10^{-4}$  implies a lifetime of only  $\sim 10^5$  yr, which is much shorter than the apparent 40-Myr cosmic-ray-exposure age<sup>5</sup> of SiC.

**Heating in the Solar System.** Some SiC was destroyed in the Solar System by the enigmatic 'chondrule-forming process' that flash-heated part of the dust to melting or even vaporization temperatures, producing chondrules, metal grains and Ca, Al-rich inclusions<sup>23</sup>. SiC would be irreversibly destroyed in the O-rich gas (C/O = 0.42), to an extent indicated by the fraction of such coarse-grained material. But the C2 chondrites from which our SiC comes contain only 50% coarse-grained material,

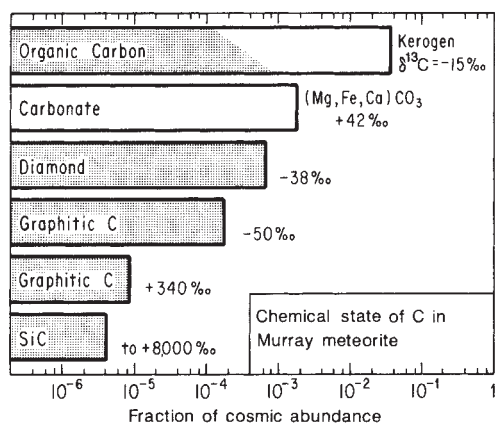


FIG. 5 Chemical state of C in primitive meteorites. Species classified as interstellar on the basis of isotope anomalies are shaded. Diamond, not graphitic C, is the dominant form of elemental C. Total organic C is more abundant still, but only a minor, poorly known, fraction of this is interstellar. SiC accounts for only  $4 \times 10^{-6}$  of the total C rather than the  $\sim 0.03$  expected if  $\sim 0.3$  of the interstellar Si came from carbon stars. Apparently carbon-star material was grossly under-represented in the solar nebula.

so this process can deplete SiC by only a factor of two. Presumably the SiC deficiency was inherited from an earlier stage, by some combination of underproduction, overdestruction and local heterogeneity.

Nuth<sup>28</sup> has suggested that the chemical state of C in primitive meteorites may reflect that in the Galaxy. Specifically, he noted that graphitic C is much rarer than organic carbon, and in view of its higher heat resistance it must have been at least as rare in the interstellar matter before Solar-System reprocessing. Apparently graphitic C, although the thermodynamically stable form of solid C in carbon-star atmospheres, is fairly rare in our Galaxy.

Following Tang and Anders<sup>2</sup>, we shall re-examine this conclusion and its underlying assumptions, on the basis of the new data in Fig. 5. Organic C still is the principal form, but judging from its modest deuterium enrichment (a factor of 2–8 greater

than terrestrial ocean water) and the rarity of D-rich 'hotspots' in Semarkona, the most D-rich chondrite<sup>29</sup>, only a minor part ( $\sim 1\%$ ?) seems to be well-preserved interstellar material. Most of the organic carbon was altered in the Solar System or formed anew by grain catalysis<sup>30,31</sup> from CO. The CO may have been in part interstellar, and in part the degradation product of reprocessed interstellar organic matter.

The new data (Fig. 5) confirm that graphitic carbon is fairly rare, actually being less abundant than diamond. There are two reasons, however, why this distribution cannot be a quantitative representation of the galactic carbon inventory.

First, a substantial part of the original interstellar material was reprocessed in the solar nebula. Although the matrix of primitive chondrites ( $\sim 50\%$  for C2 chondrites) escaped heating to  $> 500$  K, it did experience some reheating, as well as alteration by liquid water in the meteorite parent bodies. This left the sturdier forms of carbon unaffected, but volatilized or altered the organic carbon. Counteracting these destructive processes, some organic and graphitic C was formed metastably from CO in the solar nebula<sup>30</sup>. Thus meteoritic abundances of organic and graphitic carbon cannot be related directly to interstellar abundances of these species.

The second consideration is the rarity of carbon-star material. We have argued above that SiC is underabundant in meteorites by a factor of  $\sim 10^{-4}$  relative to the amount injected into the interstellar medium by carbon-rich stars, implying underproduction or overdestruction of carbon-star material in protosolar matter. But diamond and graphitic C also come from carbon-rich stars (not necessarily of the same class), and so they too may be underabundant in meteorites. (Alternatively, diamond could be overabundant if it came from a single nearby supernova or red giant that contributed short-lived radionuclides.) Thus even the surviving interstellar component in meteorites may not be representative of average interstellar matter.

In conclusion, we point out the high degree of observational selection in our data. The interstellar grain types isolated thus far, diamond and SiC, are exceptionally resistant to destruction in nature and in the laboratory. Less-resistant types would have been destroyed by our isolation procedures, if they were not previously destroyed by thermal or aqueous alteration in the meteorite parent bodies. In future, more comprehensive work may give a broader representation of interstellar dust. □

Received 16 January; accepted 20 April 1989.

1. Tang, M., Lewis, R. S. & Anders, E. *Meteoritics* **22**, 462–463 (1987).
2. Tang, M. & Anders, E. *Geochim. cosmochim. Acta* **52**, 1235–1244 (1988).
3. Bernatowicz, T. et al. *Nature* **330**, 728–730 (1987).
4. Zinner, E., Tang, M. & Anders, E. *Nature* **330**, 730–732 (1987).
5. Tang, M. & Anders, E. *Astrophys. J.* **335**, L31–L34 (1988).
6. Tang, M., Anders, E. & Zinner, E. *Lunar planet. Sci.* **19**, 1177–1178 (1988).
7. Clayton, D. D. *Astrophys. J.* **344**, 191–195 (1988).
8. Seab, C. G. in *Interstellar Processes* (eds Hollenbach, D. J. & Thronson, H. A. Jr) 491–512 (Reidel, Dordrecht, 1987).
9. Lambert, D. L., Gustafsson, B., Eriksson, K. & Hinkle, K. H. *Astrophys. J. Suppl. Ser.* **62**, 373–425 (1986).
10. Wallerstein, G. *Science* **240**, 1743–1750 (1988).
11. Larimer, J. W. & Bartholomay, M. *Geochim. cosmochim. Acta* **43**, 1455–1466 (1979).
12. Ash, R. D., Arden, J. W., Grady, M. M., Wright, I. P. & Pillinger, C. T. *Nature* **336**, 228–230 (1988).
13. Wright, I. P., Ash, R. D., Grady, M. M., Pillinger, C. T. & Tang, M. *Meteoritics* **23**, 312–313 (1988).
14. Hawkins, I. & Jura, M. *Astrophys. J.* **317**, 926–950 (1987).
15. Lambert, D. L. in *Cosmic Abundances of Matter* (ed. Waddington, C. J.) 168–199 (American Institute of Physics, New York, 1989).
16. Schramm, D. N. & Olive, K. A. *Ann. N.Y. Acad. Sci.* **395**, 236–241 (1982).
17. Clayton, D. D. & Hoyle, F. *Astrophys. J.* **203**, 490–496 (1976).
18. Arnould, M. *Phil. Trans. R. Soc. A* **323**, 251–267 (1987).
19. Iben, I. & Renzini, A. *Rev. Astr. Astrophys.* **21**, 271–342 (1983).

20. Tielens, A. G. G. M. in *Carbon in the Galaxy: Studies from Earth and Space* (ed. Tarter, J.) (NASA CP, in the press).
21. Arnould, M. & Beelen, W. *Astr. Astrophys.* **33**, 215–230 (1974).
22. Woosley, S. E. & Hoffmann, R. D. in *Nucleosynthesis and Chemical Evolution* (eds Hauck, B., Maeder, A. & Meynet, G.) (Saas-Fee, 1986).
23. Wood, J. A. *A. Rev. Earth planet. Sci.* **16**, 53–72 (1988).
24. Reeves, H. *Astrophys. J.* **231**, 229–235 (1979).
25. Wasserburg, G. J. in *Protostars and Planets II* (eds Black, D. C. & Matthews, M. S.) 703–737 (Univ. Arizona Press, Tucson, 1985).
26. Cameron, A. G. W. *Icarus* **60**, 416–427 (1984).
27. Lewis, R. S., Tang, M., Wacker, J. F., Anders, E. & Steel, E. *Nature* **326**, 160–162 (1987).
28. Nuth, J. A. *Nature* **318**, 166–168 (1985).
29. Walker, R. M. & Zinner, E. *Meteoritics* **23**, 307 (1988).
30. Hayatsu, R. & Anders, E. *Topics curr. Chem.* **99**, 1–37 (1981).
31. Lewis, J. S. & Prinn, R. G. *Astrophys. J.* **238**, 357–364 (1980).
32. Smith, V. V. & Lambert, D. L. *Astrophys. J.* **294**, 326–328 (1985).
33. Smith, V. V. & Lambert, D. L. *Astrophys. J.* **311**, 843–863 (1986).
34. Harris, M. J., Lambert, D. L. & Smith, V. V. *Astrophys. J.* **292**, 620–627 (1985).
35. Harris, M. J., Lambert, D. L. & Smith, V. V. *Astrophys. J.* **299**, 375–385 (1985).
36. Dominy, J. F., Wallerstein, G. & Suntzeff, N. B. *Astrophys. J.* **300**, 325–338 (1986).
37. Dominy, J. F. & Wallerstein, G. *Astrophys. J.* **317**, 810–818 (1987).

ACKNOWLEDGEMENTS. We thank J. A. Nuth, III, for a review, and D. D. Clayton, B. T. Draine, D. L. Lambert, D. N. Schramm and G. Wallerstein for advice. This work was supported in part by NASA (E.A.) and the NSF (E.Z.). E.Z. also thanks R. M. Walker for his support.



# A fusion protein required for vesicle-mediated transport in both mammalian cells and yeast

Duncan W. Wilson\*, Celeste A. Wilcox†, Gregory C. Flynn\*, Ellson Chen‡, Wun-Jing Kuang‡, William J. Henzel‡, Marc R. Block†§, Axel Ullrich‡§ & James E. Rothman\*

\* Department of Biology, Princeton University, Princeton, New Jersey 08540, USA

† Department of Biochemistry, Stanford University, Stanford, California 94305, USA

‡ Genentech Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

A protein sensitive to *N*-ethylmaleimide catalyses the fusion of transport vesicles with Golgi cisternae in a mammalian cell-free system. By cloning and sequencing its gene from Chinese hamster ovary cells and by use of *in vitro* assays, we show that this fusion protein is equivalent to the *SEC18* gene product of the yeast *Saccharomyces cerevisiae*, known to be essential for vesicle-mediated transport from the endoplasmic reticulum to the Golgi apparatus. The mechanism of vesicular fusion is thus highly conserved, both between species and at different stages of transport.

PROTEIN transport between membrane-bound compartments is accomplished by the budding and fusion of transport vesicles<sup>1</sup>. Previous studies have shown that the molecular basis of this process can be investigated by reconstituting the transport of a glycoprotein encoded by vesicular stomatitis virus (VSV) between successive cisternae of the Golgi apparatus<sup>2,3</sup>. This cell-free transport is mediated by a population of non-clathrin-coated vesicles which, after budding from donor cisternae, attach to acceptor cisternae, where they are uncoated and then fuse<sup>4,5</sup>. In addition to Golgi membranes, the transport requires both cytosol and ATP, and is stimulated by fatty acyl coenzyme A (CoA)<sup>6</sup>.

Fusion, which involves the stepwise maturation of the vesicle-cisterna junction through a series of intermediates, requires several protein factors, one of which is inhibited by the sulphhydryl reagent *N*-ethylmaleimide (NEM) under mild conditions (1 mM for 15 min at 0°C)<sup>6,7</sup>. That this factor is needed for membrane fusion was indicated by the observation that inactivation with NEM did not block budding, but led to the accumulation of uncoated, unfused vesicles on the surface of acceptor cisternae<sup>8</sup>. An NEM-sensitive fusion protein, termed NSF, was thus purified on the basis of its ability to restore transport to NEM-inactivated Golgi membranes<sup>9</sup>. The protein consists of four identical subunits of relative molecular mass ( $M_r$ ) 76,000 (76K) and is found in both membrane-bound and soluble forms. In the presence of ATP, which stabilizes the activity of the soluble form of NSF, the protein is released from Golgi membranes<sup>9</sup>.

Here we report the primary sequence of NSF from Chinese hamster ovary (CHO) cells: the sequence is very similar to that reported recently<sup>10</sup> for the product of the *SEC18* gene of the yeast *Saccharomyces cerevisiae*, known to be essential for transport from the endoplasmic reticulum (ER) to Golgi *in vivo*. We

also show that the *SEC18* gene encodes NSF activity in yeast cytosol and in accompanying papers report that in addition to its role within the Golgi stack, NSF is required for export from the ER in animal cells<sup>11</sup> and during endocytosis<sup>12</sup>.

## Sequencing the gene encoding NSF

As the N-terminus of NSF was found to be blocked, the purified protein<sup>9</sup> was cleaved with cyanogen bromide and the isolated peptides subjected to partial amino-acid sequence analysis<sup>13</sup>. Table 1 shows the resulting peptide sequences used to design synthetic oligonucleotide probes<sup>14</sup> for the identification of NSF clones within a CHO cell  $\lambda$ gt-10 complementary (c) DNA library<sup>15</sup>. Of the 68 positive clones obtained on screening  $1 \times 10^6$  recombinant  $\lambda$ -phage, 11 contained inserts of ~4 kilobases (kb), equivalent to the size of an RNA molecule detected when using the same oligonucleotides to probe a northern blot of CHO cell messenger (m) RNA (data not shown). Nucleotide sequence analysis of the largest clone (Fig. 1) revealed an open reading frame of 752 residues with the potential to encode a protein of  $M_r$  83K. This is in close agreement with the apparent molecular mass of NSF ( $M_r$  76K)<sup>9</sup> determined by SDS polyacrylamide gel electrophoresis (SDS PAGE).

The precise nature of the N-terminus of NSF remains uncertain: an ATG codon marked with asterisks (Fig. 1) could serve as the initiation codon, however as the reading frame is open upstream of this codon it is possible that translation is initiated further upstream, beyond the boundary of the clone. Nevertheless, because the ATG codon within the clone is in an appropriate context for translational initiation<sup>16</sup>, an A residue occurs at the -3 position (with respect to the first base of the triplet) while G occurs at the +4 and -6 positions, it is reasonable to suppose that this triplet acts as the initiation codon, giving rise to a protein of 744 residues.

TABLE 1 Sequence analysis of NSF-derived peptides

| Peptide number | Peptide sequence determined from pure NSF | Location within NSF |
|----------------|-------------------------------------------|---------------------|
| 1              | DPSILKGEPASGKRQKIEVGLVV                   | 164-186             |
| 2              | GIXGLDKFSDIFRRASFVRFPXEIVE                | 227-254             |
| 3              | XNAREPKVVGPEILNKYVGES                     | 285-306             |
| 4              | EIGLPDEKGRQLIHHTARM                       | 402-422             |
| 5              | RGHQLLSADVDIKELAVETKNFS                   | 423-445             |
| 6              | EKAESLQVTXGDFLASLENDIKPAFGTNQEDYASY       | 476-510             |
| 7              | DPEYRVRKFLALLREEGASPLDF                   | 729-751             |

Partial sequence analysis of NSF-derived peptides, numbered as in Fig. 1 and corresponding to residues within the predicted amino-acid sequence of NSF (Fig. 1), as indicated. Amino acids not determined are denoted by X. Automated Edman degradation was performed with a model 470A Applied Biosystems gas phase sequencer equipped with a 120A PTH amino-acid analyser. PTH amino acids were integrated with a Nelson analytical model 3000 data system. Sequence interpretation was performed on a VAX 11/875 Digital Equipment Corporation computer.

§ Present addresses: Laboratoire de Physiologie Cellulaire, Université Joseph Fourier, B.P. 53X, 38041 Grenoble CEDEX, France. (M.R.B.); Max-Planck Institut für Biochemie, 8033 Martinsried, FRG (A.U.).

|        |                                                                |      |                                                                    |     |  |
|--------|----------------------------------------------------------------|------|--------------------------------------------------------------------|-----|--|
| *** 10 |                                                                | 480  |                                                                    | 490 |  |
| 1      | TCG GAA CCG GAC GTG TCC GCG AAG ATG GCG GGC CGG AGT ATG        | 1429 | AAA GCA GAG AGT CTG CAG GTG ACA AGA GGA GAC TTC CTG GCT            |     |  |
|        | Ser Glu Pro Asp Val Ser Ala Lys Met Ala Gly Arg Ser Met        |      | <u>Lys Ala Glu Ser Leu Gln Val Thr Arg Gly Asp Phe Leu Ala</u>     |     |  |
| 43     | CAA GCT GCG AGG TGC CCT ACC GAT GAA TTG TCC TTA AGC AAC        | 1471 | TCT TTG GAG AAT GAC ATC AAA CCA GCT TTT GGC ACA AAC CAA            |     |  |
|        | Gln Ala Ala Arg Cys Pro Thr Asp Glu Leu Ser Leu Ser Asn        |      | <u>Ser Leu Glu Asn Asp Ile Lys Pro Ala Phe Gly Thr Asn Gln</u>     |     |  |
| 85     | TGT GCA GTT GTG AGC GAG AAG GAT TAC CAG TCT GGA CAG CAT        | 1513 | GAC GAT TAT GCC AGT TAC ATT ATG AAT GGA ATC ATC AAG TGC            |     |  |
|        | Cys Ala Val Val Ser Glu Lys Asp Tyr Gln Ser Gly Gln His        |      | <u>Glu Asp Tyr Ala Ser Tyr Ile Met Asn Gly Ile Ile Lys Trp</u>     |     |  |
| 127    | GTG ATT GTG AGG ACC TCC CCC AAT CAC AAA TAC ATA TTT ACA        | 1555 | GGT GAC CCT GTG ACC CCT GTC CTG GAT GAC GGA GAG CTG CTG            |     |  |
|        | Val Ile Val Arg Thr Ser Pro Thr Asp Glu Leu Ser Lys Thr        |      | <u>Gly Asp Pro Val Thr Arg Val Leu Asp Asp Gly Glu Leu Leu</u>     |     |  |
| 169    | CTG AGG ACC CAT CGC TCG GTG GTT CCC GGA AGC GTT GCA TTC        | 1597 | GTC CAG CAG ACT AAA AAC AGT GAC CGT ACG CCT CTG GTT AGC            |     |  |
|        | Leu Arg Thr His Pro Ser Val Val Pro Gly Ser Val Ala Phe        |      | <u>Val Gln Gln Thr Lys Asn Ser Asp Arg Thr Pro Leu Val Ser</u>     |     |  |
| 211    | AGC TTA CCT CAG CGA AAA TGG CCT GGA CTT TCT ATT GGC CAG        | 1639 | GTC CTT CTG GAA GGC CCC CCT CAC AGC GGC AAG ACT GCC CTA            |     |  |
|        | Ser Leu Pro Gln Arg Lys Trp Ala Gly Leu Ser Ile Gly Gln        |      | <u>Val Leu Leu Glu Gly Pro Pro His Ser Gly Lys Thr Ala Leu</u>     |     |  |
| 253    | GAA ATA GAA GTC GCC TTG TAT TCA TTT GAC AAA GCC AAG CAA        | 1681 | GCT GCA AAG ATC GCA GAG GAA TCC AAC TTT CCC TTC ATC AAG            |     |  |
|        | Glu Ile Glu Val Ala Leu Tyr Ser Phe Asp Lys Ala Lys Gln        |      | <u>Ala Ala Lys Ile Ala Met Lys Lys Ile Phe Asp Asp Ala Tyr Lys</u> |     |  |
| 295    | TGC ATC GGC ACC ATG ACC ATC GAG ATT GAC TTT CTG CAG AAA        | 1723 | ATC TGC TCC CGC GAT AAG ATG ATT GGC TTT TCT GAG ACA GCC            |     |  |
|        | Cys Ile Gly Thr Met Thr Ile Glu Ile Asp Phe Leu Gln Lys        |      | <u>Ile Cys Ser Pro Asp Lys Met Ile Gly Phe Ser Glu Thr Ala</u>     |     |  |
| 337    | AAA AAC ATC GAC TCC AAC CCA TAT GAT ACC GAC AAG ATG GCA        | 1765 | AAG TGT CAG GCC ATG AAG AAG ATC TTT GAT GAT GCT TAC AAA            |     |  |
|        | Lys Asn Ile Asp Ser Asn Pro Tyr Asp Thr Asp Lys Lys Phe Ala    |      | <u>Lys Cys Gln Ala Met Lys Lys Ile Phe Asp Asp Ala Tyr Lys</u>     |     |  |
| 379    | GCA GAA TTC ATC CAG CAG TTC AAC AAC CAG GCC TTC TCA GTG        | 1807 | TCT CAA CTC AGT TGT GTG GTG GTG GAT GAC ATC GAA AGG CTA            |     |  |
|        | Ala Glu Phe Ile Gln Gln Phe Asn Asn Gln Ala Phe Ser Val        |      | <u>Ser Gln Leu Ser Cys Val Val Val Asp Asp Ile Glu Arg Leu</u>     |     |  |
| 421    | GGA CAG CAG CTT GTC TTT AGC TTT AAT GAT AAG CTC TTC GGA        | 1849 | CTT GAT TAT GTC CCC ATT GGC CGC CGG TTC TCC AAC CTG CTG            |     |  |
|        | Gly Gln Gln Leu Val Phe Ser Phe Asn Asp Lys Leu Phe Gly        |      | <u>Leu Asp Tyr Val Pro Ile Gly Pro Arg Phe Ser Asn Leu Val</u>     |     |  |
| 463    | TTA CTG GTG AAG GAC ATT GAA GCC ATG GAC CCT AGC ATC CTG        | 1891 | TTG CAA GCT CTT CTT GTG TTA CTA AAG AAG GCA CCT CCT CAG            |     |  |
|        | Leu Leu Val Lys Asp Ile Glu Ala Met <u>Asp Pro Ser Ile Leu</u> |      | <u>Leu Gln Ala Leu Leu Val Leu Lys Lys Ala Pro Pro Gln</u>         |     |  |
| 505    | AAG GGA CAG CCT GCA TCA GGC AAA AGG CAG AAG ATT GAA CTC        | 1933 | GCC GGC AAG CTT CTC ATC ATC GGC ACC ACC AGC GGC AAA GAT            |     |  |
|        | <u>Lys Gly Glu Pro Ala Ser Gly Lys Arg Gln Lys Ile Glu Val</u> |      | <u>Gly Arg Lys Leu Leu Ile Ile Gly Thr Thr Ser Arg Lys Asp</u>     |     |  |
| 547    | GGA CTG GTT GTT GGA AAT AGT CAA GTT GCA TTT GAA AAA GCA        | 1975 | CTC CTT CAG GAA ATG GAG ATG CTG AAT GCC TTC ACC ACC ACC            |     |  |
|        | <u>Gly Leu Val Val</u> Gly Asn Ser Gln Val Ala Phe Lys Lys Ala |      | <u>Val Leu Gln Glu Met Lys Lys Ile Phe Asp Ala Phe Ser Thr Thr</u> |     |  |
| 58     | GAA AAT TCA TCG CTC AAT CTT ATT GGC AAA GCT AAA ACC AAG        | 2017 | ATC CAC GTG CCC AAC ATT GCC ACA GGA GAG CAG CTG CTG GAA            |     |  |
|        | Glu Asn Ser Ser Leu Asn Leu Ile Gly Lys Ala Lys Thr Lys        |      | <u>Ile His Val Pro Asn Ile Ala Thr Gly Glu Gln Leu Leu Glu</u>     |     |  |
| 631    | GAA AAT GCG CAG TCA ATC ATC AAT CCT GAC TGG AAC TTT GAG        | 2059 | GCT TTG GAG CTT CTG GGC AAC TTC AAA GAT AAG GAA CGC ACC            |     |  |
|        | Glu Asn Arg Gln Ser Ile Ile Asn Pro Asp Trp Asn Phe Glu        |      | <u>Ala Leu Glu Leu Leu Gly Asn Phe Lys Asp Lys Glu Arg Thr</u>     |     |  |
| 673    | AAA ATG GGA ATC GGC GGT CTG GAT AAG GAG TTT TCA GAT ATT        | 2101 | ACA ATT GCT CAG CAA CTC AAA GGG AAG AAG CTC TGC ATA GGA            |     |  |
|        | Lys Met <u>Gly Ile Gly Gly Leu Asp Lys Glu Phe Ser Asp Ile</u> |      | <u>Thr Ile Ala Gln Gln Val Lys Gly Lys Lys Val Trp Ile Gly</u>     |     |  |
| 715    | TTC CGA CGA GCA TTT GCC TCT CGG GTG TTC CCT CCG GAG ATT        | 2143 | ATC AAG AAG TTA CTA ATG TTT ATT GAG ATG TCC CTG CAG ATG            |     |  |
|        | <u>Phe Arg Arg Ala Phe Ala Ser Arg Val Phe Pro Pro Glu Ile</u> |      | <u>Ile Lys Lys Leu Leu Met Leu Ile Glu Met Ser Leu Gln Met</u>     |     |  |
| 757    | GTG GAG CAG ATG GGT TGT AAA CAT GTT AAA GGC ATC CTG TTG        | 2185 | GAC CCT GAG TAT CGG GTG AGG AAA TTC TTG GCC CTC TTG AGA            |     |  |
|        | <u>Val Glu</u> Gln Met Gly Cys Lys His Val Lys Gly Ile Leu Leu |      | <u>Asp Pro Glu Tyr Arg Val Arg Lys Phe Leu Ala Leu Leu Arg</u>     |     |  |
| 799    | TAC GGC CCC CCT GGT TGT GGT AAG ACT CTC TTG GCT CGA CAG        | 2227 | GAA GAA GGC GCT AGC CCC CTG GAC TTT GAT TGA                        |     |  |
|        | Tyr Gly Pro Pro Gly Cys Gly Lys Thr Leu Leu Ala Arg Gln        |      | <u>Glu Glu Gly Ala Ser Pro Leu Asp Phe Asp</u>                     |     |  |
| 841    | ATT GGC AAG ATC CTG AAT GCG AGA GAA CCC AAG GTG GTC AAT        |      |                                                                    |     |  |
|        | Ile Gly Lys Met <u>Leu Asn Ala Arg Glu Pro Lys Val Val Asn</u> |      |                                                                    |     |  |
| 883    | GGG CCA GAA ATC CTT AAC AAA TAT GTG GGA GAA TCG GAG GCG        |      |                                                                    |     |  |
|        | <u>Gly Pro Glu Ile Leu Asn Lys Tyr Val Gly Glu Ser</u> Glu Ala |      |                                                                    |     |  |
| 925    | AAC ATT CGT AAA CTT TTT GCT GAT GCC GAA GAG GAG CAA AGG        |      |                                                                    |     |  |
|        | Asn Ile Arg Lys Leu Phe Ala Asp Ala Glu Glu Gln Arg            |      |                                                                    |     |  |
| 967    | AGG CTT GGT GCT AAC AGT GCG TTA CAC ATC ATC TTT GAT            |      |                                                                    |     |  |
|        | Arg Leu Gly Ala Asn Ser Gly Leu His Ile Ile Phe Asp            |      |                                                                    |     |  |
| 1009   | GAA ATC GAT GCC ATC TGC AAG CAG AGA GGC AGC ATG GCT GGC        |      |                                                                    |     |  |
|        | Glu Ile Asp Ala Ile Cys Lys Gln Arg Gly Ser Met Ala Gly        |      |                                                                    |     |  |
| 1051   | AGC ACC GGA GTT CAG GAC ACC GTT GTC AAC CAG CTG TCC            |      |                                                                    |     |  |
|        | Ser Thr Gly Val His Asp Thr Val Val Asn Gln Leu Leu Ser        |      |                                                                    |     |  |
| 1093   | AAG ATT GAT GGC GTT CAG CAG CTG AAC AAC ATC CTG GTT ATT        |      |                                                                    |     |  |
|        | Lys Ile Asp Gly Val Glu Gln Leu Asn Asn Ile Leu Val Ile        |      |                                                                    |     |  |
| 1135   | GGG ATG ACC AAT AGA CCA GAT TTG ATA GAT GAG GCT CTC CTT        |      |                                                                    |     |  |
|        | Gly Met Thr Asn Arg Pro Asp Thr Lys Leu Ile Asp Glu Ala Leu    |      |                                                                    |     |  |
| 1177   | CGA CTT GGA AGA CTG GAA GTT AAA ATG GAG ATA GGC TTG CCA        |      |                                                                    |     |  |
|        | Arg Pro Gly Arg Leu Glu Val Lys Met <u>Glu Ile Gly Leu Pro</u> |      |                                                                    |     |  |
| 1219   | GAT GAG AAG GGT CGA CTG CAG ATC CTT CAT ATC CAC ACA GCA        |      |                                                                    |     |  |
|        | <u>Asp Glu Lys Gly Val Glu Lys Ile Asn His Ile His Thr Ala</u> |      |                                                                    |     |  |
| 1261   | AGG ATG AGA GGA CAT CAG TTA CTG TCT GCA GAT GTG GAC ATT        |      |                                                                    |     |  |
|        | <u>Arg Met Arg Gly His Gln Leu Leu Ser Ala Asp Val Asp Ile</u> |      |                                                                    |     |  |
| 1303   | AAG GAA CTG GCC GTG GAG ACT AAG AAT TTC AGT GGC GCT GAG        |      |                                                                    |     |  |
|        | <u>Lys Glu Leu Val Glu Thr Lys Phe Ser</u> Gly Ala Glu         |      |                                                                    |     |  |
| 1345   | CTG GAA GGT CTG GTG CGA GCA GCA CAG TCG ACG GCC ATG AAC        |      |                                                                    |     |  |
|        | Leu Glu Gly Leu Val Arg Ala Ala Gln Ser Thr Ala Met Asn        |      |                                                                    |     |  |
| 1387   | CGA CAC ATA AAG GCC AGT ACG AAA CTG GAA GTA GAC ATG GAG        |      |                                                                    |     |  |
|        | Arg His Ile Lys Ala Ser Thr Lys Val Glu Val Asp Met <u>Glu</u> |      |                                                                    |     |  |

FIG. 1 Nucleotide sequence and deduced amino-acid sequence of NSF from CHO cells. Nucleotides are numbered on the left; amino acids are numbered over the nucleotide sequence, starting at the first amino acid of the reading frame. Regions corresponding to peptide sequences derived from purified NSF are underlined and numbered in brackets according to Table 1. Synthetic oligonucleotide probes of the following sequences were used to screen for positive cDNA clones, and correspond to each of the peptides as indicated in the following (for peptide 6, two overlapping oligonucleotides were prepared). Peptide 3 probe: CTGCCCCACATACTTGTCAGGATCTCAGGGCCATTC-ACCACCTTGGGCTCCCGGGGATT. Peptide 4 probe: CATCCGGCTGTGTGG-ATGTGCAGGATCTGCAGCCGGCCCTTCTCATCAGGCAGGCCAATCTC. Peptide 5 probe: GAAGTTCTTGGTCTCCACAGCCAGCTCCTTGATGTCCACATCAGCAGAC-AGCAGCTGGTGGCC. Peptide 6 probe A: GTAGTCCTCCTGGTGGTCCAAA-GGCAGCTTGATGTCTATCTCCAGGGAGGCCAGGAA. Peptide 6 probe B: ATT-CTCCAGGGAGGCCAGGAAGTCGCCATTGGTCACCTGCAGGGACTCAGCCTTCTC-CAT.

METHODS. Synthetic probes were radiolabelled by 5'-terminal phosphorylation using T4 polynucleotide kinase (PL biochemicals) and [ $\gamma$ - $^{32}$ P]ATP (Amersham,  $>5,000$  Ci  $\text{mM}^{-1}$ ). Unincorporated radionucleotides were removed by gel filtration on a Sepharose G-50 column. A  $\lambda$ gt-10 cDNA library was prepared using poly A<sup>+</sup> RNA from CHO cells as previously described<sup>28</sup>, and was screened using a pool of the five radiolabelled oligonucleotides. Hybridization of the nitrocellulose-bound phage was performed at 42 °C for 16 h at low stringency: 30% formamide,  $5 \times \text{SSC}$  ( $1 \times \text{SSC} = 150$  mM NaCl, 15 mM Na<sub>3</sub> Citrate); filters were then subjected to three 10-min washes in  $0.2 \times \text{SSC}$  at 42 °C. Nucleotide sequence analysis was carried out in M13-based cloning vectors<sup>29,30</sup>.



As NSF is required for membrane fusion, we sought to determine whether the NSF polypeptide contains any long stretches of hydrophobic residues, resembling the N-terminus of influenza hemagglutinin HA2 (ref. 17), which by analogy might facilitate interaction of NSF with membranes. A hydropathy analysis, however, failed to reveal such regions (data not shown). It is not unprecedented for fusogenic proteins to lack such hydrophobic segments; the envelope glycoprotein of VSV, for example, does not have a translocated hydrophobic domain<sup>17</sup>. Thus, whether NSF is actually a 'fusogen' rather than a 'fusion protein' necessary but not sufficient for fusion, remains to be determined.

### Similarity between NSF and SEC18 gene product

There is a striking similarity between NSF and the predicted product of the *SEC18* gene<sup>10,18,19</sup> of *S. cerevisiae* (Fig. 2). The two proteins are 48% identical, with an additional 15% of the residues being conservatively substituted over 757 residues for Sec18p and 752 residues for NSF. Given the evolutionary distance between yeast and animals, this suggests that the proteins have similar, if not identical, functions. When switched from growth at 25 °C to 37 °C (the restrictive temperature) strains bearing a temperature-sensitive mutation within *SEC18* accumulate ER that contains cell-surface and vacuolar proteins in incompletely glycosylated forms<sup>10,18,19</sup>; Sec18p is thus essential for the process of vesicular transport from the ER to the Golgi apparatus.

Expression of the *SEC18* gene *in vitro* and *in vivo* leads to production of two forms of Sec18p, differing in  $M_r$  by ~2K (ref. 10). This may reflect the presence of two in-frame initiation codons<sup>10</sup>, and could have a bearing on the finding that the NSF reading frame is open upstream of an ATG codon lying in an appropriate context for translational initiation (see above). Following SDS PAGE<sup>9</sup>, however, purified NSF and NSF immunoprecipitated from a CHO cell lysate both appeared as a single molecular species.

### Detection of NSF activity in yeast cytosol

The sequence homology between NSF and Sec18p suggests they might have similar biochemical properties; the question arises therefore as to whether Sec18p performs an NSF-like function in yeast. A more preliminary question, however, is whether NSF activity can be detected in yeast. Although yeast cytosol has previously been shown to contain soluble factors able to drive inter-cisternal transport *in vitro*<sup>20</sup>, under the assay conditions used NSF was supplied in excess by CHO Golgi membranes. Our results (Fig. 3) show that yeast cytosol does indeed contain NSF activity that restores transport to NEM-inactivated CHO Golgi. Moreover, this activity shares several of the distinctive features of CHO cell NSF. First, pretreatment of cytosol with 5mM NEM for 15 min at 0 °C inactivates yeast NSF activity (although it should be noted that CHO NSF is inactivated by only 1 mM NEM under these conditions<sup>6</sup>). Second, that yeast NSF is rapidly inactivated in the absence of ATP is consistent with the known effect of ATP on the stability of CHO NSF<sup>9</sup>. Note that both NSF and Sec18p possess a consensus sequence for ATP binding<sup>21</sup> (Fig. 2).

To determine whether the *SEC18* gene product is responsible for the NSF activity of yeast cytosol, an *S. cerevisiae* strain was transformed with a multicopy plasmid bearing the wild-type *SEC18* gene. The presence of this plasmid has been shown to lead to overproduction of Sec18p (ref. 10). Cytosol was prepared from transformed and untransformed strains and then assayed for NSF activity. The level of NSF activity in the transformed strain was at least 10-fold higher than that of the untransformed strain at the lowest concentration of protein assayed (Fig. 4). This activity, resulting from Sec18p, is NEM-sensitive (Fig. 4) and requires ATP for stability (Fig. 5) as observed for the yeast factor in cytosol from untransformed strains.

The question of whether the NSF activity in yeast cytosol is

provided by Sec18p alone, or by a family of NSF-like factors functioning at different points within the secretory pathway was also addressed. Cytosol was prepared from a yeast strain bearing the temperature-sensitive allele *sec18-1* and then assayed for NSF activity at both permissive (25 °C) and restrictive (37 °C) temperatures. No NSF activity was detected at either temperature (Fig. 6), showing that Sec18p is inactive *in vitro* even at 25 °C. Lack of NSF activity within this mutant cytosol suggests that Sec18p alone is responsible for the NSF activity of yeast

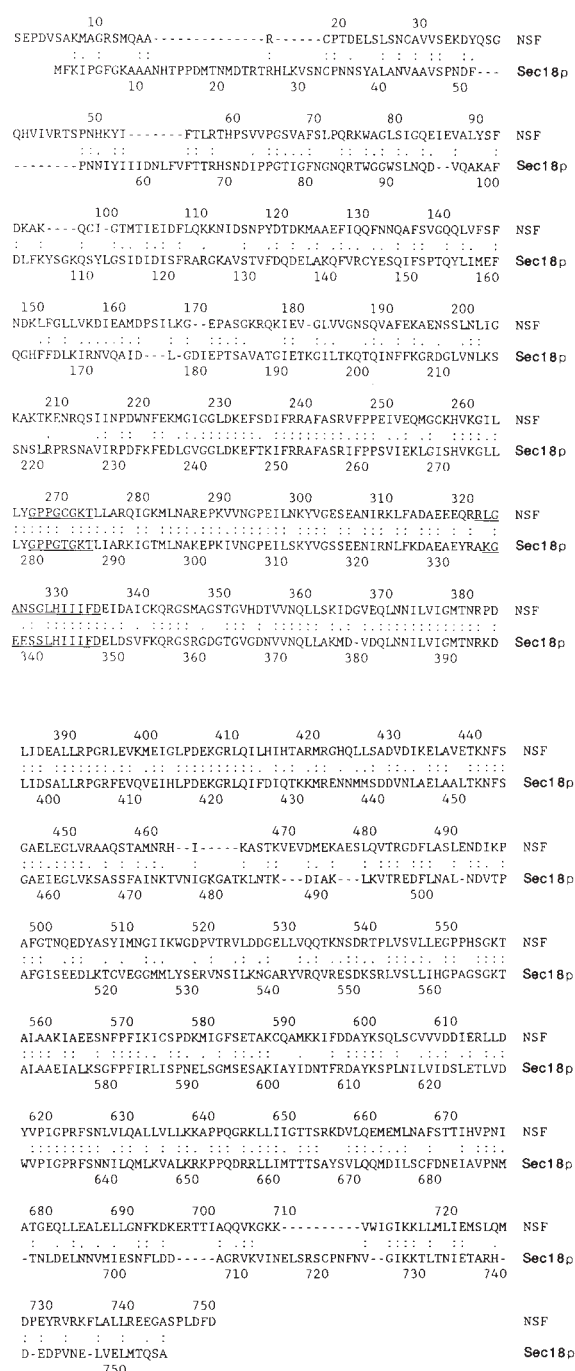


FIG. 2 NSF shares extensive sequence similarity with the product of the *SEC18* gene. Amino-acid residues in both Sec18p and NSF are numbered as shown. Aligned residues identical in NSF and Sec18p are denoted by (:) whereas (.) marks similar residues. Gaps, inserted to optimize alignment, are shown by a dash. The underlined residues correspond to the two consensus regions A (NSF residues 268–275) and B (NSF residues 323–336) constituting a putative ATP binding site.

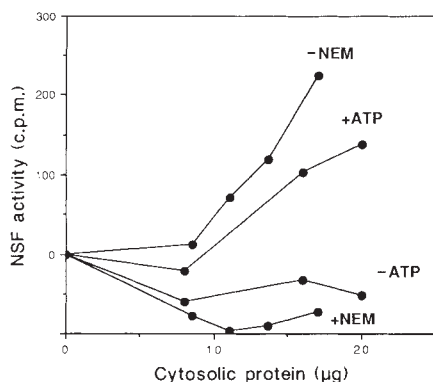


FIG. 3 Yeast cytosol contains an NEM-sensitive, ATP-stabilized factor with NSF-activity (c.p.m., counts per minute). Cytosol was prepared from *S. cerevisiae* SEY2102.1 (*ade2-101 ura3-52 leu2-3, 112 suc2-Δ9 gal2 PEP4::LEU2*), depleted of dithiothreitol (DTT) by passage over a G25M gel filtration column (as described below), and incubated at 0 °C in the presence (+NEM) or absence (–NEM) of 5 mM NEM. After 15 min, 6 mM DTT was added to the +NEM reaction and a mixture of DTT and NEM added to the –NEM reaction, to final concentrations of 6 mM and 5 mM, respectively. Increasing amounts of this mixture were then added to an NSF-dependent Golgi transport assay. In a separate experiment, two identical preparations of SEY2102.1 cytosol were desalted over a G25M column (at 0 °C) in either the presence (+ATP) or absence (–ATP) of 5 mM ATP. ATP was added back to the –ATP sample (to a final concentration of 5 mM) after filtration and both G25 cytosol preparations assayed for NSF activity. Background incorporation of UDP-[<sup>3</sup>H]N-acetylglucosamine into VSV-G protein (in the absence of cytosol) was 300 c.p.m., and was subtracted from the data.

**METHODS.** To prepare yeast extract for NSF assay, 1.5 l of minimal media (6.7 g l<sup>-1</sup> yeast nitrogen base, 20 g l<sup>-1</sup> dextrose, amino acids and nucleotides as required) were inoculated and grown to OD<sub>660</sub> 0.7–1.5 at either 30 °C (SEY2102.1) or 25 °C (*sec18-1* mutant). Untransformed strains were grown in minimal media to facilitate comparison with transformed strains, grown in selective media to ensure plasmid maintenance. NSF activity was similarly detectable in cytosol prepared from untransformed strains grown on rich media (data not shown). Where appropriate, an aliquot of culture was retained for estimation of the proportion of plasmid-containing cells in the population (typically 50–90% of cells contained plasmids at the time of cytosol preparation). Cells were pelleted, washed and resuspended (2 ml buffer per g wet weight of cells) in breaking buffer; 100 mM HEPES.KOH (pH 7.8), 5 mM ATP, 1 mM EGTA, 500 mM KCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 2 μg ml<sup>-1</sup> aprotinin, 10 μg ml<sup>-1</sup> leupeptin, 2 μM pepstatin, 1 mM phenylmethylsulphonyl fluoride (PMSF) and 1 mM *o*-phenanthroline, then disrupted using a bead-beater (Biospec products). To obtain active cytosol it was necessary to fill the bead beater chamber to capacity, avoiding air bubbles. Beating was for three periods of 20 s with intervals of 1 min, and the chamber maintained at 0 °C within an ice and salt-water jacket throughout the procedure. Debris was removed by pelleting at 16,000 r.p.m. in a Sorvall SA600 rotor for 30 min and the supernatant clarified at 70,000 r.p.m. in a Beckman TLA 100.3 ultracentrifuge rotor for 30 min. The resulting high speed supernatant, termed 'cytosol', was either frozen in liquid nitrogen and stored at –80 °C, or passed over a 9.5-ml Sephadex (Pharmacia) G25M gel filtration column equilibrated with NSF buffer; 50 mM HEPES.KOH (pH 7.2), 50 mM KCl, 5 mM ATP, 5 mM MgCl<sub>2</sub>, 2 mM DTT, 200 μM PMSF, 1% polyethylene glycol 4000 at 0 °C. Aliquots of the void fraction (termed G25 cytosol) were frozen in liquid nitrogen and stored at –80 °C. G25 cytosol depleted of DTT (for NEM-sensitivity studies) or ATP was prepared by omitting either component from the column buffer. Whenever two or more samples were to be compared, aliquots of cytosol were simultaneously thawed, G25-filtered and corrected to the same protein concentration (typically 5–8 mg ml<sup>-1</sup>) before simultaneous assay. NSF-dependent transport assays were carried out in 25-μl reaction volumes containing 2.5 μl each of NEM-treated donor and acceptor membranes, 2.5 μl of NSF-free CHO-15B cytosol<sup>9</sup>, 10 μM palmitoyl-CoA, 50 μM ATP, 2 mM creatine phosphate, creatine kinase (7.3 I.U. ml<sup>-1</sup>), 250 μM UTP, 0.4 μM UDP-[<sup>3</sup>H]N-acetylglucosamine (0.5 μCi; 1 Ci = 37 GBq) in an assay buffer containing 25 mM HEPES.KOH (pH 7.0), 15 mM KCl, 2.5 mM Mg(OAc)<sub>2</sub> and 0.2 M sucrose (from the Golgi membrane fractions). The yeast cytosol which was to be tested for NSF activity (up to 5 μl) was added last, and thawed immediately before use. Assays were maintained at 37 °C for 1 h, unless otherwise stated, and VSV-G protein immunoprecipitated as described<sup>9</sup> except that incubation was at room temperature for 15 min. All assays were performed in duplicate and the mean value plotted.

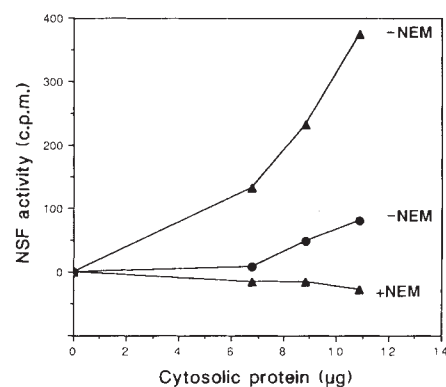


FIG. 4 Increasing the amount of yeast Sec18p *in vivo* results in increased levels of NEM-sensitive yeast NSF activity *in vitro*: circles, G25 cytosol from *S. cerevisiae* SEY2102.1; triangles, SEY2102.1 transformed with a multicopy plasmid bearing the *SEC18* gene. G25 cytosols were treated with 5 mM NEM (+NEM) or not (–NEM) as described in Fig. 3 legend. Background incorporation (245 c.p.m.) was subtracted from the data.

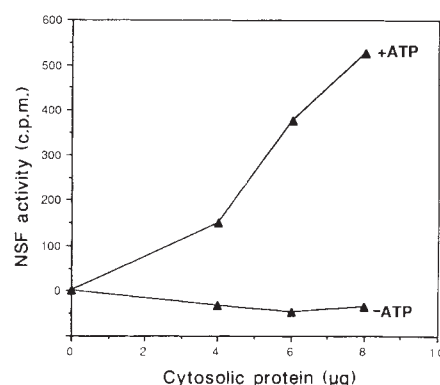


FIG. 5 Sec18p-stimulated NSF activity demonstrates dependence on ATP for stability. G25 cytosol was prepared from SEY2102.1 transformed with a multicopy plasmid bearing the *SEC18* gene. ATP was depleted (–ATP) or not (+ATP) as described in Fig. 3 legend, then each sample frozen immediately. Samples were then thawed and assayed simultaneously. Background incorporation (305 c.p.m.) was subtracted from the data.

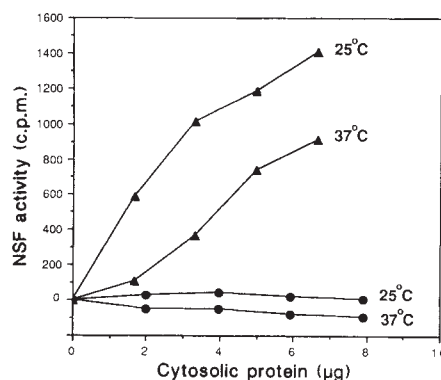


FIG. 6 G25 cytosol prepared from a *sec18-1* mutant shows no NSF activity at permissive (25 °C) and non-permissive (37 °C) temperatures unless the strain is transformed with a *SEC18*-containing plasmid: circles, untransformed *sec18-1* mutant; triangles, transformed *sec18-1* mutant. Assays were carried out under standard conditions (37 °C for 1 h) or at 25 °C for 2 h as indicated. Background incorporation (330 c.p.m., 37 °C incubation; 170 c.p.m., 25 °C incubation) were subtracted from the data.



cytosol. Introduction of the wild-type *SEC18* gene to the *sec18-1* mutant restored NSF activity (Fig. 6).

## Discussion

The *SEC18* gene in yeast encodes what seems to be the functional equivalent of NSF, a protein first defined according to its requirement for vesicular transport in a mammalian cell-free system<sup>6-9</sup>. Over-expression of the *SEC18* gene enhanced NSF activity in yeast cytosol, whereas a lesion within *SEC18* abolished all NSF activity. An alternative explanation of these results is that Sec18p stimulates the expression or activity of yeast NSF. We believe, however, that this is extremely unlikely because of the extensive sequence similarity between yeast Sec18p and mammalian NSF.

It is surprising that NSF, purified on the basis of its activity in promoting inter-cisternal transport in Golgi stacks, should be homologous to Sec18p, defined by its role in protein export from the ER. The functional equivalence of CHO NSF and yeast Sec18p in our cell-free Golgi-transport system leads to a simple explanation of the phenotype of *sec18* mutants: NSF is required at multiple stages of transport, so that its inactivation *in vivo* results in the accumulation of vesicles at the first fusion step, blocking transport from the ER to Golgi. A strong prediction, tested in the accompanying paper<sup>11</sup>, is that NSF should also be required for ER-to-Golgi transport in animal cells. Another prediction is that ER-derived transport vesicles (in addition to ER cisternae) should accumulate in *sec18* temperature-sensitive mutants at the restrictive temperature, because of the inactivation of vesicle fusion. Indications of this have been reported<sup>18</sup>.

Our findings provide a strong link between the biochemical results with animal cell systems and genetic results with yeast. That NSF activity is absent *in vitro* in extracts from a yeast mutant defective in transport *in vivo*, implies that NSF plays a

role in the cell concordant with expectations from studies in cell-free systems. Moreover, although the function performed by Sec18p was previously unknown it is now clear that *SEC18* encodes a protein essential for fusion. It is reasonable therefore to anticipate that functions for other *SEC* genes will be similarly assigned. There are at least three other mammalian protein factors that participate in the NSF-dependent fusion pathway (a soluble NSF attachment protein<sup>22</sup>, an integral membrane receptor for NSF<sup>22</sup> and soluble factor B<sup>23</sup>), all of which can now be expected to have yeast homologues defective in fusion of ER-derived transport vesicles.

How general is the role of NSF in intracellular membrane fusion events? NSF is required for transport into *trans* Golgi cisternae *in vitro*<sup>24</sup> and *sec18-1* mutants have been reported to be defective in endocytosis<sup>25</sup>. In an accompanying paper, we show that NSF is required for fusion in the pathway of endocytosis in animal cells *in vitro*<sup>12</sup>. In addition, reassembly of the nuclear envelope *in vitro*<sup>26</sup> requires an NEM-sensitive factor (W. Dunphy and J. Newport, personal communication) as does the formation of polygonal ER networks in association with microtubules<sup>27</sup>. In these latter cases, however, the NEM-sensitive component has yet to be identified.

It seems that a pool of soluble proteins that includes NSF provides a common reservoir of components to catalyse fusion at multiple subcellular locations. It is attractive to envisage the same 'fusion machine'<sup>18</sup> assembling from such dispersed subunits in response to the formation of a suitable scaffold at a vesicle-target membrane junction, whenever and wherever one forms. Specificity would be built into the vesicle-targeting process and not into the fusion machinery. Assembly of the fusion machinery in response to the completion of targeting, and dispersal during or following fusion, would prevent spurious fusion events from occurring and enable the components to be recycled for later rounds. □

Received 23 March; accepted 14 April 1989.

1. Palade, G. *Science* **109**, 347-358 (1975).
2. Balch, W. E., Dunphy, W. G., Braell, W. A. & Rothman, J. E. *Cell* **39**, 405-416 (1984).
3. Braell, W. A., Balch, W. E., Dobbertin, D. C. & Rothman, J. E. *Cell* **39**, 511-524 (1984).
4. Orci, L., Glick, B. S. & Rothman, J. E. *Cell* **46**, 171-184 (1986).
5. Orci, L., Malhotra, V., Amherdt, M., Serafini, T. & Rothman, J. E. *Cell* **56**, 357-368 (1989).
6. Glick, B. S. & Rothman, J. E. *Nature* **326**, 309-312 (1987).
7. Balch, W. E., Glick, B. S. & Rothman, J. E. *Cell* **39**, 525-536 (1984).
8. Malhotra, V., Orci, L., Glick, B. S., Block, M. R. & Rothman, J. E. *Cell* **54**, 221-227 (1988).
9. Block, M. R., Glick, B. S., Wilcox, C. A., Wieland, F. T. & Rothman, J. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7852-7856 (1988).
10. Eakle, K. A., Bernstein, M. & Emr, S. D. *Mol. cell. Biol.* **8**, 4098-4109 (1988).
11. Beckers, C. J. M., Block, M. R., Glick, B. S., Rothman, J. E. & Balch, W. E. *Nature* **339**, 397-398 (1989).
12. Diaz, R., Mayorga, L. S., Weidman, P. J., Rothman, J. E. & Stahl, P. D. *Nature* **339**, 398-400 (1989).
13. Henzel, W. J., Rodriguez, H. & Watanabe, C. *J. Chromatogr.* **404**, 41-52 (1987).
14. Lathe, R. *J. Molec. Biol.* **183**, 1-12 (1985).
15. Huyhn, T., Young, R. & Davis, R. in *Practical Approaches in Biochemistry* (ed. Grover, D.) (IRL, Oxford, 1984).
16. Kozak, M. *J. molec. Biol.* **196**, 947-950 (1987).
17. Stegmann, T., Doms, R. W. & Helenius, A. *Rev. Biophys. Biophys. Chem.* (in the press).

18. Novick, P., Field, C. & Schekman, R. *Cell* **21**, 205-215 (1980).
19. Novick, P., Ferro, S. & Schekman, R. *Cell* **25**, 461-469 (1981).
20. Dunphy, W. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 1622-1626 (1986).
21. Chin, D. T., Goff, S. A., Webster, T., Smith, T. & Goldberg, A. L. *J. biol. Chem.* **263**, 11718-11728 (1988).
22. Weidman, P. J., Melancon, P., Block, M. R. & Rothman, J. E. *J. cell. Biol.* **108**, 1589-1596 (1989).
23. Wattenberg, B. W. & Rothman, J. E. *J. biol. Chem.* **261**, 2208-2213 (1986).
24. Rothman, J. E. *J. biol. Chem.* **262**, 12502-12510 (1987).
25. Riezman, H. *Cell* **40**, 1001-1009 (1985).
26. Dunphy, W. G. & Newport, J. W. *J. cell. Biol.* **106**, 2047-2056 (1988).
27. Dabora, S. L. & Sheetz, M. P. *Cell* **54**, 27-35 (1988).
28. Yarden, Y., *et al. Nature* **323**, 226-232 (1986).
29. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
30. Messing, J., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9**, 309-321 (1981).

ACKNOWLEDGEMENTS. We thank Robert Fuller for helpful discussions. The multicopy yeast plasmid bearing the *SEC18* gene was provided by Kurt Eakle and Scott Emr. *S. cerevisiae* strains were provided by Scott Emr, Chris Kaiser and Randy Schekman. This work was supported by the National Institutes of Health and an American Cancer Society grant to J.E.R. G.C.F. is a Damon Runyon postdoctoral fellow. M.R.B. was a senior fellow of the American Cancer Society (California Branch).

# Universality in colloid aggregation

M. Y. Lin<sup>\*†</sup>, H. M. Lindsay<sup>†</sup>, D. A. Weitz<sup>\*‡</sup>, R. C. Ball<sup>‡</sup>,  
R. Klein<sup>§</sup> & P. Meakin<sup>||</sup>

<sup>\*</sup> Exxon Research and Engineering Co., Route 22E Annandale,  
New Jersey 08801, USA

<sup>†</sup> Physics Department, Emory University, Atlanta, Georgia 30322, USA

<sup>‡</sup> The Cavendish Laboratory, Madingley Road, Cambridge CB3 9HE, UK

<sup>§</sup> Fakultät für Physik, Universität Konstanz, Konstanz, FRG

<sup>||</sup> E. I. DuPont de Nemours Co., Experimental Station, Wilmington,  
Delaware 19880-0356, USA

**THE aggregation of colloidal particles is of fundamental importance in colloid science and its applications. The recent application of scaling concepts<sup>1,2</sup> has resulted in a much deeper understanding of the structure of colloidal aggregates and the kinetics of their formation. Two distinct, limiting regimes of irreversible colloid aggregation have been identified<sup>3</sup>. Diffusion-limited colloid aggregation occurs when there is negligible repulsive force between the colloidal particles, so that the aggregation rate is limited solely by the time taken for clusters to encounter each other by diffusion. Reaction-limited colloid aggregation occurs when there is still a substantial, but not insurmountable, repulsive force between the particles, so that the aggregation rate is limited by the time taken for two clusters to overcome this repulsive barrier by thermal activation. These regimes correspond to the limiting cases of rapid and slow colloid aggregation that have long been recognized in colloid science<sup>4</sup>. An intriguing possibility suggested by recent work is that each of these limiting regimes of colloid aggregation is universal, independent of the chemical details of the particular colloid system. Here we investigate the aggregation of three chemically different colloidal systems under both diffusion-limited and reaction-limited aggregation conditions. A scaling analysis of light-scattering data is used to compare the behaviour and provides convincing experimental evidence that the two regimes of aggregation are indeed universal.**

The two regimes have been studied theoretically<sup>5-8</sup>, by computer simulation<sup>9-14</sup> and by experimental studies of various colloids<sup>3,15-21</sup>. Each regime is distinguished by several distinct and characteristic features. The structures of the aggregates in both regimes are fractal, so that the mass (or number of particles in the cluster) scales as  $M \propto (R_g/a)^{d_f}$ , where  $R_g$  is the radius of gyration of the cluster and  $a$  is the radius of its constituent particles. The fractal dimension is  $d_f \approx 1.8$  for diffusion-limited colloid aggregation (DLCA) and  $d_f \approx 2.1$  for reaction-limited colloid aggregation (RLCA). For both regimes, the cluster mass distribution,  $N(M)$ , exhibits dynamic scaling, in that the shape becomes constant and universal in time. For DLCA,  $N(M)$  is slightly peaked around the average cluster mass,  $\bar{M}$ , falling exponentially beyond it. For RLCA, on the other hand, the cluster mass distribution has a power-law form, again with an exponential cutoff,  $N(M) \sim \bar{M}^{\tau-2} M^{-\tau} \exp(-M/\bar{M})$ . In both cases, only  $\bar{M}$  depends on time: for DLCA,  $\bar{M}$  grows linearly in time, whereas for RLCA,  $\bar{M}$  grows exponentially.

Here we investigate three very different colloids: colloidal gold, colloidal silica and polystyrene latex. The colloidal gold has primary particles with  $a = 7.5$  nm and an initial volume fraction of  $\phi_0 \approx 3 \times 10^{-6}$ . The colloid is initially charge-stabilized by citrate ions adsorbed on the gold particle surfaces. Aggregation is initiated by the addition of pyridine, a neutral molecule which displaces the adsorbed citrate ions and thereby reduces the stabilizing charge. The aggregation rate is controlled by the amount of pyridine added: the final pyridine concentration is  $10^{-2}$  M for DLCA and  $10^{-4}$  M for RLCA. The interparticle bonds formed are probably metallic. The colloidal silica has  $a = 3.5$  nm and is initially charge-stabilized by  $\text{OH}^-$  or  $\text{SiO}^-$  on

the surface. Aggregation is induced by addition of NaCl, which decreases the Debye-Hückel screening length, thereby decreasing the repulsive barrier between the particles. For DLCA,  $\phi_0 \approx 2 \times 10^{-6}$  and the salt concentration is 1.7 M; for RLCA,  $\phi_0 \approx 1 \times 10^{-5}$  and the salt concentration is 0.6 M. The pH is maintained at  $\geq 11$  by addition of NaOH, which ensures that siloxane bonds are formed upon aggregation. The polystyrene latex has  $a = 19$  nm and is initially charge-stabilized by adsorbed carboxylic acid groups. For DLCA,  $\phi_0 \approx 8 \times 10^{-7}$  and HCl is added to a concentration of 1.2 M, both neutralizing the surface charge and decreasing the screening length. For RLCA,  $\phi_0 \approx 7 \times 10^{-6}$  and NaCl is added to a concentration of 0.2 M, decreasing the screening length. The particle surfaces deform on bonding, leading to large van der Waals forces between the particles.

Graphic suggestion of the universal behaviour can be seen in the transmission electron micrographs shown in Fig. 1. Typical clusters formed in both the DLCA and RLCA regimes of aggregation are shown for each of the three colloids, and for computer-simulated clusters. The DLCA clusters are all more open and tenuous, reflecting their lower fractal dimension ( $d_f < 2$ ), whereas the RLCA clusters are more compact, having  $d_f > 2$ . However, the similarity between the structures of the clusters within each regime is striking, suggesting a universal nature for the aggregation processes that lead to their formation.

To compare critically the behaviour of the colloids in each regime, we use static light scattering to measure the fractal

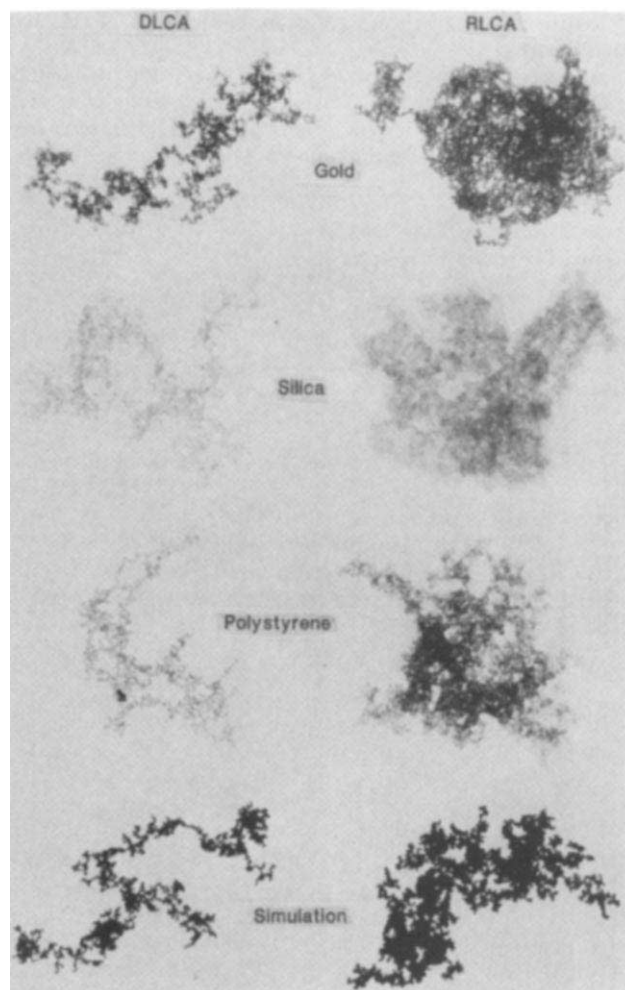


FIG. 1 Transmission electron micrographs of typical clusters of gold, silica and polystyrene colloids prepared by both diffusion-limited and reaction-limited cluster aggregation, and computer simulation. Note the striking similarity in the structure of the clusters of different colloids in each regime.

<sup>†</sup> Present address: Department of Physics, Princeton University, Princeton, New Jersey 08544, USA.

<sup>\*</sup> To whom correspondence should be addressed.



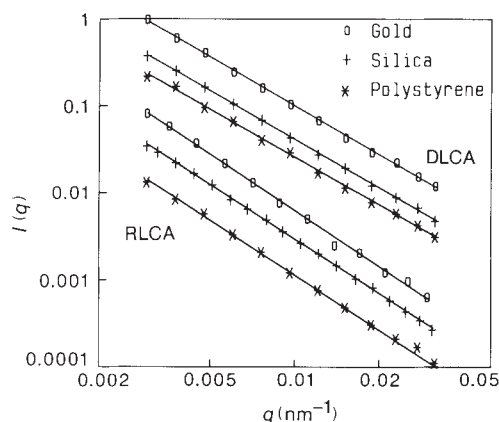


FIG. 2 Static light scattering measurements, showing universal behaviour in each regime. The linear behaviour of the data in the logarithmic plot is indicative of fractal clusters. For DLCA (top), the slopes give  $d_f=1.86$  for gold,  $d_f=1.85$  for silica and  $d_f=1.82$  for polystyrene. For RLCA (bottom) the slopes are consistently higher, with  $d_f=2.14, 2.07$  and  $2.09$ , respectively.

structure of the aggregates, and quasielastic light scattering (QELS) to measure the aggregation kinetics and to probe the shape of the cluster mass distributions. We show that, for each regime, the QELS data can be scaled onto a single master curve, the shape of which depends on the key features of the aggregation process. Comparison of the shapes of the master curves for different colloids provides a critical test of the universality of the aggregation process in each regime.

The sensitivity of light scattering to the properties of colloidal aggregates arises from the dependence of the scattered intensity on the cluster size and the scattering vector,  $q = (2\pi n/\lambda) \sin \theta/2$ . Here  $\lambda = 488$  nm is the wavelength of the laser light used,  $n$  is the index of refraction of the fluid and  $\theta$  is the scattering angle. Because the fractal clusters are self-similar in structure, the scattered intensity is a function of the product  $qR_g$ . At low  $qR_g$ , the internal structure of the aggregate is not resolved, and the scattering is completely coherent, so that the intensity scales as  $M^2$ , independent of  $q$ . At high  $qR_g$ , however, the fractal structure of the cluster is resolved and the scattering intensity reflects this with a  $q^{-d_f}$  dependence. In this case the total intensity scales linearly with  $M$ .

Both static and dynamic light scattering average over the cluster mass distribution,  $N(M)$ . Static light scattering measures the time-averaged intensity, and hence is simply a sum over the distribution, weighted by the average scattering intensity for each cluster,

$$I(q) = \frac{\sum N(M)I(qR_g)}{\sum N(M)M} \quad (1)$$

If  $q\bar{R} \gg 1$ , where  $\bar{R}$  is the average cluster size, then  $I(q) \propto q^{-d_f}$ , directly reflecting the fractal dimension of the clusters. In Fig. 2, we show the static scattering from all the colloids, obtained with  $q\bar{R} \gg 1$ . For each regime, the slopes of the data are the same for each colloid, confirming the identical fractal structure of the aggregates, with  $d_f=1.85 \pm 0.1$  for DLCA and  $d_f=2.10 \pm 0.1$  for RLCA.

By contrast to the static light scattering, QELS measures the temporal fluctuations of the scattered intensity from the aggregates resulting from their diffusive motion. We measure the autocorrelation function of these fluctuations, and here restrict our attention to its initial logarithmic slope, or 'first cumulant',  $\Gamma_1$ . For a single cluster,  $\Gamma_1 = q^2 D_{\text{eff}}(qR_g)$ , where the effective diffusion coefficient  $D_{\text{eff}}$  reflects the contribution of both translational and rotational motion<sup>22</sup>. When  $qR_g \ll 1$ , only translational diffusion contributes and  $D_{\text{eff}}(qR_g) = D = \zeta/R_H$ , where  $\zeta = kT/6\pi\eta$ , and  $\eta$  is the fluid viscosity. The cluster's hydrodynamic radius  $R_H$  is proportional to its radius of gyration<sup>23-25</sup>:  $R_H = \beta R_g$ , where  $\beta \approx 1$ . However, at  $qR_g \gg 1$ , rotational diffusion also contributes to the intensity fluctuations, and  $D_{\text{eff}}(qR_g)$  increases towards a limiting value of  $\sim 2D$ . This  $qR_g$  dependence of  $D_{\text{eff}}$  reflects the anisotropy of the aggregates, giving QELS an additional sensitivity to the cluster structure.

The average  $D_{\text{eff}}$  measured by QELS is a sum over the distribution, weighted by the scattering intensity<sup>22</sup>:

$$\bar{D}_{\text{eff}} = \frac{\bar{\Gamma}_1}{q^2} = \frac{\sum N(M)I(qR_g)D_{\text{eff}}(qR_g)}{\sum N(M)I(qR_g)} \quad (2)$$

This measures a different moment of the cluster mass distribution than does the static intensity. In the limit of  $q \rightarrow 0$ ,  $\bar{D}_{\text{eff}} = \bar{D}$ , providing a good measure of the average cluster size,  $\bar{R} = \zeta/\bar{D}$ .

The combination of the sensitivity to the cluster mass distribution and to rotational diffusion leads to a pronounced  $q$  dependence in the measured  $\bar{D}_{\text{eff}}$ , and provides a very sensitive probe of the aggregation process. However, to fully explore this  $q$  dependence at a single point in time during the aggregation process would require an experimentally inaccessible range of scattering angles. Instead, we exploit the dynamic scaling of the cluster mass distribution to measure  $\bar{D}_{\text{eff}}$  over a much wider range of  $q\bar{R}$ . Thus, we determine  $\bar{D}_{\text{eff}}$  over the experimentally accessible range of  $q$ , and repeat the measurements during the aggregation process, as  $\bar{R}$  increases while the shape of the cluster mass distribution remains unchanged. The values measured at each  $q$  are interpolated to obtain a series of data sets, each consisting of  $\bar{D}_{\text{eff}}(q)$  evaluated at the same time. We normalize  $\bar{D}_{\text{eff}}$  by  $\bar{D}$ , and plot the data as a function of  $q\bar{R}$ , where the required parameter,  $\bar{D} = \zeta/\bar{R}$ , for each set is determined empiri-

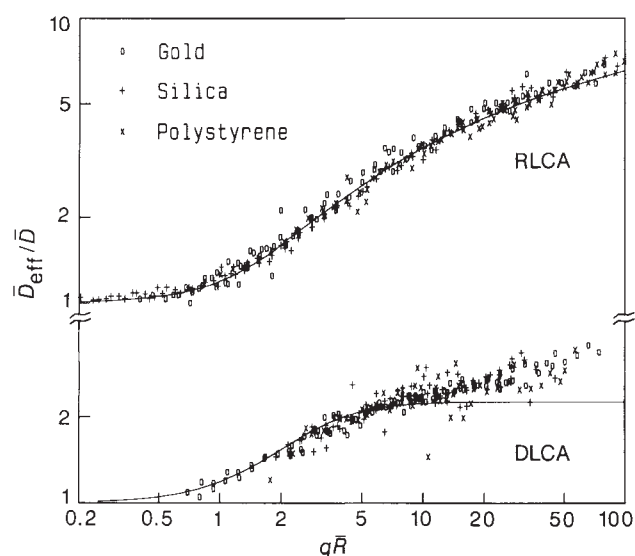


FIG. 3 Master curves from dynamic light scattering for each colloid for RLCA (top) and DLCA (bottom). All material parameters have been scaled out: the shape of the master curve is sensitive to the cluster mass distribution, the static scattering from the aggregates and the anisotropy in their structure. In each regime the master curves for different colloids are indistinguishable, showing the universality of colloid aggregation. The solid lines are theoretical calculations.

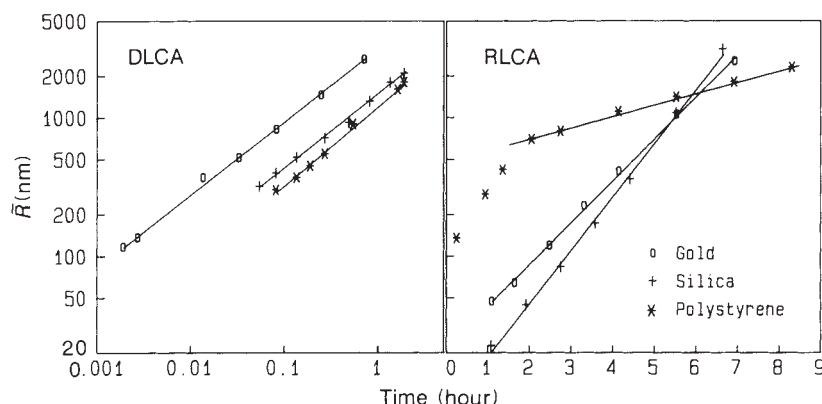


FIG. 4 Aggregation kinetics. For DLCA (left) the linear behaviour on the logarithmic plot indicates power-law kinetics,  $\bar{R} \propto t^\alpha$ . The fitted slopes give  $\alpha = 0.53$  for gold,  $\alpha = 0.54$  for silica and  $\alpha = 0.56$  for polystyrene. For RLCA (right) the linear behaviour on the semi-logarithmic plot indicates exponential kinetics. The deviation from this behaviour for polystyrene is discussed in the text.

cally by scaling the data onto a single master curve. With sufficient data, there is always a substantial overlap between data from different sets, making the scaling unambiguous. All material parameters are scaled out, so that we obtain a critical comparison between the behaviour of completely different colloids.

The master curves obtained for each colloid in each regime are plotted in Fig. 3. The shape of the master curve for DLCA is different from that of RLCA, reflecting the different shapes of  $N(M)$  for each regime; the power-law form for RLCA leads to a considerably stronger  $q$  dependence of the master curve. In each regime, the master curves for the three colloids are indistinguishable. We emphasize that the master curves for each colloid are obtained independently, and there is no free parameter in comparing them. This is striking evidence of the universality of each of the regimes of colloid aggregation.

We can also calculate<sup>26</sup> the shapes of the master curves for each regime using equation (2) and  $I(qR_g)$  from simulation clusters. The results are shown by the solid lines in Fig. 3. For DLCA, we obtain good agreement with the data, except at high  $q\bar{R}$ . For RLCA, we obtain excellent agreement at all  $q\bar{R}$ , using an exponent for the cluster mass distribution of  $\tau = 1.5$ .

The scaling values of  $\bar{R}$  accurately reflect the average cluster radius in the distribution at each time, and are therefore a useful measure of the aggregation kinetics, as shown in Fig. 4. The DLCA kinetics are power-law,  $\bar{R} \propto t^\alpha$ , for each colloid, showing linear behaviour on the logarithmic graph. The exponent is consistent with  $\alpha = 1/d_f$ , giving  $M \propto t$ , as expected<sup>7</sup>. The different offsets of the curves reflect the different primary particle radii and different initial values of  $\phi_0$ . By contrast, the RLCA kinetics are exponential, showing linear behaviour on the semi-logarithmic plot. Exponential kinetics are observed for the gold and the silica over all times, and for the polystyrene at later times.

The reaction-limited kinetics of polystyrene are initially exponential, then slow down dramatically (while still remaining exponential) after about one hour. This behaviour is observed for RLCA measurements using small polystyrene spheres, and is consistent with observations of others<sup>27</sup>. The absolute rate of aggregation, and the subsequent exponential kinetics, rule out the possibility of this representing a crossover to DLCA. As both the scaling of the master-curve data and the static scattering strongly indicate RLCA behaviour throughout, we suggest that this change results from an increase in the energy barrier between the particles at later times, associated with partial coalescence on aggregation, which has been observed for similar particles<sup>28</sup>.

We have presented experimental evidence from three different colloid systems which demonstrates that the process of colloid aggregation is universal, independent of the detailed chemical nature of the colloids. This universality can be used to help rationalize deviations that are bound to occur (for example, the polystyrene RLCA kinetics). Biological molecules exhibit markedly different aggregation behaviour<sup>27,29,30</sup>, producing

clusters with considerably higher  $d_f$ . These observations may be understood if the clusters undergo considerable restructuring during aggregation.

We conclude by recalling that it has long been recognized that the stability of lyophobic colloids can be rationalized in universal terms by considering the interaction energy between two approaching particles. What is remarkable about the results presented here is that they demonstrate that the whole process of their aggregation is universal, thereby greatly extending our ability to understand the properties of colloids and their aggregates.

*Note added in proof:* A recent paper<sup>31</sup> has suggested that the aggregation behaviour of colloidal gold is different from that of other colloids, based on an interpretation of light-scattering data which is claimed to exhibit multiple scattering resulting from the optical resonance of the metallic particles. The results presented here demonstrate conclusively that both the static and dynamic light scattering from the gold aggregates is identical to that from the more weakly scattering dielectric aggregates, providing unambiguous evidence that there is no multiple scattering from the metallic clusters, and that the aggregation of colloidal gold is the same as the other colloids in each of the limiting, universal regimes. □

Received 16 February; accepted 17 April 1989.

- Meakin, P. in *Phase Transitions* (ed. Liebowitz, J. L.) **12**, 335 (Academic, New York, 1988).
- Julien, R. & Botet, R. *Aggregation and Fractal Aggregates* (World Scientific, Singapore, 1987).
- Weitz, D. A., Huang, J. S., Lin, M. Y. & Sung, J. *Phys. Rev. Lett.* **54**, 1416–1419 (1985).
- Verwey, E. J. W. & Overbeek, J. T. G. *Theory of the Stability of Lyophobic Colloids* (Elsevier, Amsterdam, 1948).
- Cohen, R. J. & Benedek, G. B. *J. phys. Chem.* **86**, 3696–3714 (1982).
- Vicsek, T. & Family, F. *Phys. Rev. Lett.* **52**, 1669–1672 (1984).
- Van Dongen, G. J. & Ernst, M. H. *Phys. Rev. Lett.* **54**, 1396–1399 (1985).
- Ball, R. C., Weitz, D. A., Witten, T. A. & Leyvraz, F. *Phys. Rev. Lett.* **58**, 274–277 (1987).
- Meakin, P. *Phys. Rev. Lett.* **51**, 1119–1122 (1983).
- Kolb, M., Botet, R. & Julien, R. *Phys. Rev. Lett.* **51**, 1123–1126 (1983).
- Brown, W. D. & Ball, R. C. *J. Phys.* **A18**, L517–L519 (1985).
- Mountain, R. D. & Mulholland, G. W. *Langmuir* **4**, 1321–1326 (1988).
- Meakin, P., Vicsek, T. & Family, F. *Phys. Rev. B* **31**, 564–569 (1985).
- Meakin, P. & Family, F. *Phys. Rev. A* **36**, 5498–5501 (1987).
- Von Schultess, G. K., Benedek, G. B. & De Blois, R. W. *Macromolecules* **13**, 939–945 (1980).
- Weitz, D. A. & Oliveria, M. *Phys. Rev. Lett.* **52**, 1433–1436 (1984).
- Schaefer, D. W., Martin, J. E., Wiltzius, P. & Cannell, D. S. *Phys. Rev. Lett.* **52**, 2371–2374 (1984).
- Matsushita, M., Hayakawa, Y., Sumida, K. & Sawada, Y. in *Science on Form: Proceedings of the First International Conference for Science on Form* (ed. Kato, Y., Takaki, R. & Toriwaki, J.) (KTK Scientific Publishers, Tokyo, 1986).
- Aubert, C. & Cannell, D. S. *Phys. Rev. Lett.* **56**, 738–741 (1986).
- Weitz, D. A. & Lin, M. Y. *Phys. Rev. Lett.* **57**, 2037–2040 (1986).
- Martin, J. E. *Phys. Rev. A* **36**, 3415–3426 (1987).
- Lindsay, H. M., Klein, R., Weitz, D. A., Lin, M. Y. & Meakin, P. *Phys. Rev. A* **38**, 2614–2626 (1988).
- Chen, Z.-Y., Deutch, J. M. & Meakin, P. *J. chem. Phys.* **80**, 2982–2983 (1984).
- Hess, W., Frisch, H. L. & Klein, R. Z. *Phys.* **B64**, 65–68 (1986).
- Wiltzius, P. *Phys. Rev. Lett.* **58**, 710–713 (1987).
- Lin, M. Y. et al. *Proc. R. Soc. Lond. A* (in the press).
- Rarity, J. G., Seabrook, R. N. & Carr, R. G. *Proc. Roy. Soc. Lond. A* (in the press).
- Buscall, R., Mills, P. D. A., Goodwin, J. W. & Lawson, D. W. *J. chem. Soc. Faraday Trans. 1*, 4249–4260 (1988).
- Feder, J., Jossang, T. & Rosenqvist, E. *Phys. Rev. Lett.* **53**, 1403 (1984).
- Horne, D. S. *Faraday Discuss. chem. Soc.* **83**, 259–270 (1987).
- Wilcoxon, J. P., Martin, J. E. & Schaefer, D. W. *Phys. Rev. A* **39**, 2675–2688 (1989).

ACKNOWLEDGEMENTS. We thank John Dunsmair and the two Deckmans for their help with the transmission electron micrographs.



# Eutectics and the formation of amorphous alloys

R. J. Highmore & A. L. Greer

Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK

**SOLID** amorphous alloys can be made by a wide variety of techniques<sup>1</sup>. Of these, rapid liquid quenching and solid-state amorphization (SSA) in annealed multilayers can be most readily analysed according to a condensed-phase equilibrium diagram for the alloy. Here we consider how the phase equilibria in eutectic systems are affected by the interaction of the liquid components. We use such considerations to show the link between two apparently distinct types of alloy system: one capable of exhibiting SSA, in which there are compounds and a deep metastable eutectic; the other not capable of SSA, but capable of glass formation by rapid liquid quenching, in which there is only a stable eutectic. In alloys of intermediate type we propose that there exists a novel transformation, the inverse eutectic transition, which is a special case of a peritectoid transformation (two solid phases transforming to a single solid phase on cooling) in which the product phase is amorphous.

There is no known alloy system in stable equilibrium in which the liquid is preserved down to absolute zero, nor in which the glass transition temperature  $T_g$  exceeds at any composition the liquidus temperature  $T_l$ . In rapid liquid quenching, amorphous alloys are formed by limiting the time available for crystallization while cooling through the temperature interval from  $T_l$  (below which crystallization is thermodynamically possible) down to  $T_g$  (below which the system is configurationally frozen), and this is easiest at deep eutectics<sup>2,3</sup>, as has been clearly demonstrated in many systems (Fig. 1a). A deep eutectic is associated with high stability of the liquid and with ordering, which can be detected by structural measurements but which is also indicated by the liquid thermodynamics. The specific heat of the liquid exceeds, by an amount  $\Delta C_p$ , that of the mixture of crystalline elements, and correspondingly, the amounts  $\Delta H$  and  $\Delta S$  by which the enthalpy and entropy of the liquid exceed those of the elemental mixture decrease as the temperature is lowered. The effect on the enthalpy is readily observed: although the enthalpy of formation of intermetallic compounds from the elements is only weakly temperature-dependent, the heat of crystallization to intermetallic compounds of an amorphous alloy, measured near  $T_g$ , is typically only 50–70% of the equilibrium heat of fusion at the same composition<sup>4</sup>.

Especially strong interaction between the elements in the liquid state (that is, a highly negative heat of mixing in the liquid) appears to be necessary (though not sufficient *per se*) for SSA, in which thin layers of polycrystalline elements react, on annealing at temperatures far below the liquidus, to form a solid amorphous alloy (an example is Ni–Zr, shown in Fig. 1b)<sup>5</sup>. A necessary condition for SSA is that the anneal be performed below the  $T_g$  for the amorphous alloy in order to inhibit nucleation and growth of the equilibrium intermetallic compounds that are expected in a system with attractive interactions between the constituents. It is these high-melting compounds that prevent such a system from being a very good glass former by rapid quenching, restricting the composition range over which, for example, amorphous Ni–Zr alloys can be formed. The observation of SSA demonstrates that equilibrium compounds can be avoided, and that metastable equilibrium diagrams (shown in bold in Fig. 1b) can be relevant. The chosen metastable equilibrium diagram involves only the liquid alloy and the solid elemental phases, and it shows a deep metastable eutectic. For this eutectic,  $T_g$  must be greater than the eutectic temperature  $T_e$ . In contrast, systems such as Au–Si (Fig. 1a),

capable of being made amorphous by liquid quenching but not by SSA, have  $T_g < T_e$ .

In Ni–Zr, the amorphization reaction (involving mixing of the elements) is strongly exothermic<sup>6</sup>, although on the basis of the structural similarity to eutectic melting one would expect it to be endothermic. The negative value of  $\Delta H$  with respect to the mixture of elements, however, is a natural extension of the behaviour noted above, in which  $\Delta H$  values become less positive in undercooled glass-forming alloys. If there is a eutectic transition at all, then the entropy change on melting,  $\Delta S$ , must be positive. This can be consistent with an equilibrium transition with negative  $\Delta H$  only if the transition temperature is negative. In Ni–Zr this is reasonable, as the metastable equilibrium diagram seems to show that the liquidus lines would not meet above absolute zero.

We explore the evolution from an alloy system such as Au–Si to a system such as Ni–Zr, which occurs as the attractive interaction between the components in the liquid state is made stronger. Thermodynamically this is most readily modelled by increasing the excess specific heat,  $\Delta C_p$ , of the liquid alloy. For liquids of glass-forming, deep-eutectic composition,  $\Delta C_p$  is positive at  $T_e$ , and increases approximately linearly with undercooling until the glass transition temperature is reached<sup>7,8</sup>. For the glass it is a good approximation to take  $\Delta C_p = 0$ . For simplicity we consider a symmetrical system in which the two elements have the same melting point, the eutectic composition is equiatomic, and there is no solubility in the crystalline elements. As  $\Delta C_p \approx 0$  for pure metals<sup>9</sup>, we set  $\Delta C_p = 0$  for the alloy at the ideal melting temperature  $T_{id}$  (that is, on the linear interpolation between the elemental melting points). For a binary alloy with mole fraction  $x$  of one component, we take  $\Delta C_p = \chi x(1-x)(T_{id} - T)$  for  $T < T_{id}$  (that is, ignoring the glass transition), and we use  $\chi$  as a parameter to quantify the attractive interaction of the liquid components. The effect of increasing  $\chi$  is illustrated by the solid lines in Fig. 2, which show, at the eutectic composition, the excess  $\Delta G$  of the liquid free energy over that of the mixture of crystalline elements. As expected, the eutectic temperature  $T_e$  decreases as the excess specific heat is increased. At the same

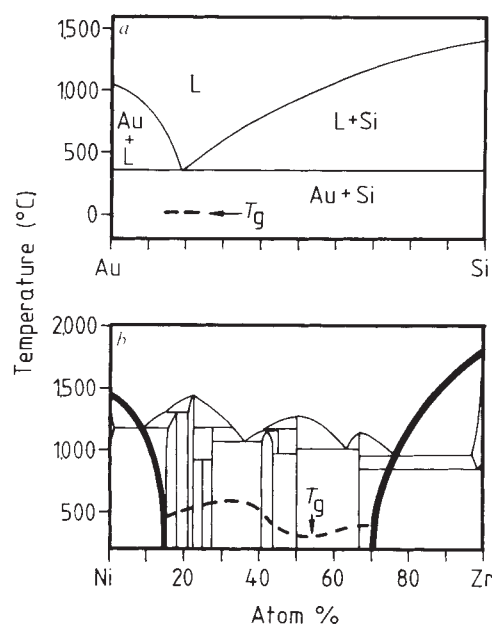


FIG. 1 a, The equilibrium phase diagram of Au–Si, the first metallic system shown to be glass-forming by rapid quenching of the liquid<sup>3</sup>. It has not been shown to be capable of amorphization by solid-state reaction of the crystalline elements.  $T_g$  values from ref. 15. b, The equilibrium phase diagram of Ni–Zr, a system exhibiting solid-state amorphization<sup>5</sup>. The bold lines are a possible metastable phase diagram in the absence of compounds, that is, for equilibria between the liquid and the elemental solid solutions only.  $T_g$  values from ref. 16.

time, another transition temperature appears and increases. This we label the inverse eutectic temperature,  $T_{ie}$ , because it is the temperature at which, on cooling, a mixture of elements would transform back to an amorphous alloy. As  $\Delta S = -(\partial\Delta G/\partial T)_P$ , the maxima of the curves in Fig. 2 show the isentropic temperature  $T_s$  (at which the extrapolated liquid entropy and crystal entropy are equal). As  $\chi$  is increased,  $T_s$  rises, and ultimately a pseudocritical point is reached at which  $T_s = T_e = T_{ie}$  and the eutectic transitions disappear.

The observed glass transition is a kinetic phenomenon caused by progressive lowering of atomic mobility as the liquid is cooled, and the temperature at which it occurs is lower for lower cooling rates. However, Kauzmann<sup>10</sup> showed that there must be a transition even for infinitely slow cooling; if such an underlying transition did not arise, the liquid would exhibit the 'apparent paradox' of having, at finite temperature, an entropy lower than that of the corresponding crystal. The isentropic temperature  $T_s$  may be regarded as an ideal glass transition for the one-component liquids studied by Kauzmann. The association of the glass transition with  $T_s$  may fail for alloys, in which the configurational entropy has not only a topological but also a chemical component. If it remains true that the glass transition occurs when the topological entropies of liquid and crystal are the same, then for an alloy system in which the liquid has a higher degree of chemical order than the crystalline phase mixture, the ideal  $T_g$  will be less than  $T_s$ . In eutectic systems, the entropy arising from the solute content of the elemental phases may outweigh the chemical mixing entropy in the liquid or amorphous phase. The liquid mixing entropy is expected to be very low in systems in which there is a strong tendency for ordering. Measurements of the entropy of mixing of the liquid components in a binary alloy are rare, but for Al-Ca the mixing entropy has been found to be negative over part of the composition range<sup>11</sup>, indicating not only a low configurational mixing entropy but also significant vibrational or electronic contributions. Further evidence that liquid alloys can have lower entropies than corresponding solid phases is provided by observations of inverse melting<sup>12</sup>. This phenomenon of melting on cooling could not occur if  $T_g \geq T_s$  because a positive or zero  $\Delta S$  would be frozen into the glass. Examples are the partial transformation of the  $\epsilon$ -Cu<sub>3</sub>Sn phase to a tin-rich liquid on cooling<sup>13</sup>, and, in the absence of partitioning, the suggested vitrification of Ti-(30 at% Cr) body-centred cubic solid solution on annealing<sup>14</sup>. We consider that in transformations such as SSA, which occur during low-temperature anneals, the temperature below which the non-crystalline phase is configura-

tionally frozen is close to the ideal glass transition temperature, and may therefore be below  $T_s$ .

The influence of the glass transition on the behaviour shown in Fig. 2 can now be examined. In the configurationally frozen state, the glass preserves the  $\Delta H$  and  $\Delta S$  values of the liquid at  $T_g$ . The excess of the free energy of the glass over that of the mixture of crystalline elements is given by the tangent to the  $\Delta G(T)$  curve at  $T_g$ . We have argued that, for simple one-component liquids,  $T_g \geq T_s$ , and that for chemically ordered liquids,  $T_g$  can be less than  $T_s$ . Accordingly the tangents in Fig. 2 are constructed at points that show the variation from  $T_g > T_s$  to  $T_g < T_s$  as  $\chi$  is increased. With this behaviour, giving a temperature dependence of  $\Delta G$  shown by the dashed lines, the inverse eutectic transition and the pseudocritical point will still be observed. Corresponding to the behaviour indicated by the dashed lines, Fig. 3a shows the evolution, as the attractive interaction in the liquid is increased, of the binary phase diagram representing the equilibria between the alloy liquid and the solid elemental phases. At one extreme, with ideal interactions in the liquid and  $\chi = 0$ , is a normal eutectic. At the other extreme, with a highly negative heat of mixing in the liquid, is a system such as Ni-Zr. In between lies a system with the pseudocritical point. The existence of this point is readily seen by examining the surface composed of the eutectic reaction tie-lines. This surface folds back at a critical value of  $\chi$ . At a value of the attractive interaction before the fold is reached, the binary phase diagram shows both eutectic and inverse eutectic transitions, as shown in Fig. 3b. The pseudocritical point is attained when the transformation temperatures  $T_e$  and  $T_{ie}$  meet. A pseudocritical point is likely in the transition from a system with  $T_g < T_e$  to one with  $T_g > T_e$ , but it is not inevitable and is avoided if the ideal glass transition temperature is always greater than  $T_s$ . In that case, the condition of equal entropy is never reached, there is no inverse melting transition and the eutectic tie-line surface does not fold back but instead slopes steeply downwards, indicating an ever-deepening eutectic. The tie-line may reach absolute zero

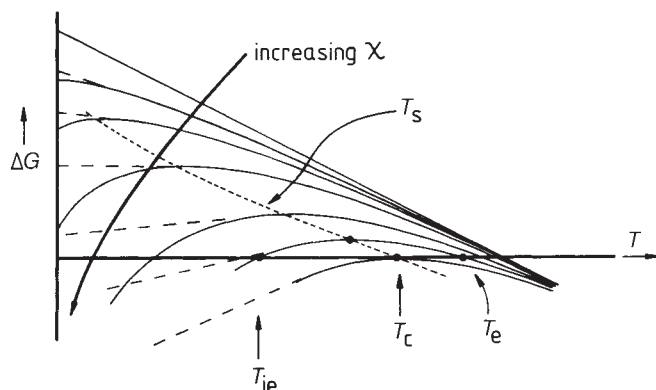


FIG. 2 The excess,  $\Delta G$ , of the alloy liquid free energy over that of the solid elements, as a function of temperature at the eutectic composition. The solid curves are calculated assuming that the excess specific heat of the liquid varies linearly with temperature and is proportional to  $\chi$  (see text for details). The dashed lines show the effect of the glass transition. On one curve are marked the eutectic temperature  $T_e$ , the inverse eutectic temperature  $T_{ie}$  and the isentropic temperature  $T_s$ . The pseudocritical point is marked  $T_c$ . The temperature scale starts above absolute zero and hence the  $\Delta G$  curves do not have zero gradient at the vertical axis.

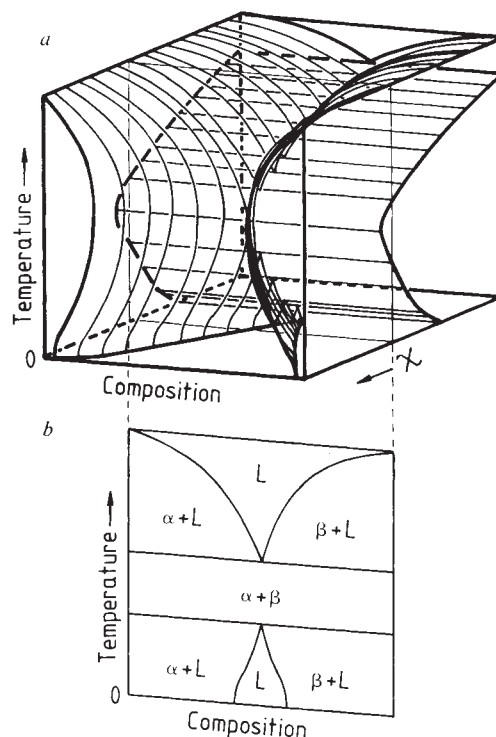


FIG. 3 a, The evolution of an idealized eutectic phase diagram (without compounds) as the degree of chemical order in the liquid, characterized by  $\chi$ , is increased. The fold in the surface of horizontal tie-lines occurs at the pseudocritical point. b, A section through a at an intermediate value of  $\chi$ , showing both eutectic and inverse eutectic transitions.

for moderate solute contents, and for higher solute contents the eutectic temperature could be regarded as negative.

Alloy systems showing SSA have a greater readiness to order than those that form glasses by rapid quenching but not by SSA. When the ordering tendency is strong enough, and in the absence of intermetallic compounds (avoidable in low-temperature anneals), the glass transition can be at higher temperature than the liquidus, and the amorphous phase may be in equilibrium even down to absolute zero. With an intermediate ordering tendency, there may be an inverse eutectic transition, affecting critical cooling rates for amorphous-phase formation by rapid quenching, and giving possibilities for SSA in systems exhibiting metastable phase diagrams such as that in Fig. 3b. □

Received 13 February; accepted 28 April 1989.

1. Johnson, W. L. *Prog. Mater. Sci.* **30**, 81–134 (1986).
2. Cohen, M. H. & Turnbull, D. *Nature* **189**, 131–132 (1961).
3. Klement, W., Willens, R. H. & Duwez, P. *Nature* **187**, 869–870 (1960).
4. Garone, E. & Battezzati, L. *Phil. Mag. B* **52**, 1033–1045 (1985).
5. Clemens, B. M., Johnson, W. L. & Schwarz, R. B. *J. Non-Cryst. Solids* **61** & **62**, 817–822 (1984).
6. Cotts, E. J., Meng, W. J. & Johnson, W. L. *Phys. Rev. Lett.* **57**, 2295–2298 (1986).
7. Evans, P. V., Garcia-Escorial, A., Donovan, P. E. & Greer, A. L. *Mater. Res. Soc. Symp. Proc.* **57**, 234–253 (1987).
8. Greer, A. L. & Evans, P. V. in *Principles of Solidification and Materials Processing* (ed. Trivedi, R.) (Trans. Tech., Zurich, in the press).
9. Perepezhko, J. H. & Paik, J. S. *J. Non-Cryst. Solids* **61** & **62**, 113–118 (1984).
10. Kauzmann, W. *Chem. Rev.* **43**, 219–256 (1948).
11. Sommer, F., Lee, J.-J. & Predel, B. *Z. Metallk.* **74**, 100–104 (1983).
12. Greer, A. L. *J. less-Common Met.* **140**, 327–334 (1988).
13. Hansen, M. & Anderko, K. *Constitution of Binary Alloys* 633–634 (McGraw-Hill, New York, 1958).
14. Blatter, A. & von Allmen, M. *Phys. Rev. Lett.* **54**, 2103–2106 (1985).
15. Chen, H. S. & Turnbull, D. *J. chem. Phys.* **44**, 2560–2571 (1968).
16. Altounian, Z., Tu Guo-hua & Strom-Olsen, J. *J. appl. Phys.* **54**, 3111–3116 (1983).

## Possible climate change due to SO<sub>2</sub>-derived cloud condensation nuclei

T. M. L. Wigley

Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK

It has been hypothesized that climate may be noticeably affected by changes in cloud condensation nuclei (CCN) concentrations, caused either by changes in the flux of dimethylsulphide (DMS) from the oceans<sup>1,2</sup> and/or by man-made increases in the flux of sulphur dioxide (SO<sub>2</sub>) into the atmosphere<sup>3</sup>. When oxidized, the sulphur compounds produce non-sea-salt sulphate (n.s.s.-SO<sub>4</sub><sup>2-</sup>) aerosols, which may act as CCNs. The CCN changes affect climate by altering the number density and size distribution of droplets in clouds, and hence their albedo. Here I am concerned primarily with the possible effects of SO<sub>2</sub>. Because the increase in SO<sub>2</sub> emissions has been largely in the Northern Hemisphere, this raises the possibility of a cooling of the Northern Hemisphere relative to the Southern<sup>3</sup>. By comparing observed differences in hemispheric-mean temperatures with results from a simple climate model, one can place limits on the possible magnitude of any SO<sub>2</sub>-derived forcing. The upper limit is sufficiently large that the effects of SO<sub>2</sub> may have significantly offset the temperature changes that have resulted from the greenhouse effect.

Schwartz<sup>3</sup> has documented the likely increase in sub-micrometre-sized n.s.s.-SO<sub>4</sub><sup>2-</sup> aerosol in the Northern Hemisphere, which he attributes largely to increased SO<sub>2</sub> fluxes. This is supported by Greenland ice-core analyses of sulphate<sup>4,5</sup>, which show a two- to threefold increase during the twentieth century, and is consistent with the estimated six- to sevenfold increase in global SO<sub>2</sub> flux this century<sup>6</sup>, more than 90% of which has been in the Northern Hemisphere. Furthermore, CCN concentrations over the north Atlantic are typically three or more times greater than those over southern oceans<sup>7</sup>. Cloud reflectivity

differences have also been observed<sup>8</sup> in the vicinity of ship tracks, which could be partly a result of SO<sub>2</sub> effects.

Twomey *et al.*<sup>7</sup> and Charlson *et al.*<sup>1</sup> have estimated the effect of CCN changes on planetary albedo and radiation balance. For the cloud type most affected, stratiform water clouds, Charlson *et al.* (Fig. 1 and Table 1 of ref. 1) give results that may be expressed as

$$\Delta\alpha = -0.067 \Delta N/N \quad (1)$$

where  $\alpha$  is the TOA (top-of-atmosphere) albedo of the cloud averaged over the solar spectrum, and  $N$  is the cloud droplet number density (equation (1) assumes constant liquid-water content). A similar result has been given by Twomey *et al.* (ref. 7, Fig. 6b). (The example given in ref. 1 gives a proportionality constant of  $-0.053$ , but I have recalculated the constant from their Fig. 1. The uncertainty is about  $\pm 20\%$ .) As  $N \propto C^m$  where  $C$  is the CCN concentration and  $m \approx 0.8$  (ref. 9), equation (1) becomes

$$\Delta\alpha = -0.053 \Delta C/C \quad (2)$$

If the albedo change occurred in all suitable clouds over the Northern Hemisphere (such clouds cover roughly 45% of the global ocean area<sup>1</sup>), then the radiative balance for the hemisphere would be perturbed by an amount

$$\Delta Q_{\text{NH}} = -4.92 \Delta C/C \text{ (W m}^{-2}\text{)} \quad (3)$$

Unfortunately, the SO<sub>2</sub>-induced change in CCN concentrations is unknown, and it is difficult even to guess at a reasonable value. Charlson *et al.*<sup>1</sup> assume  $\Delta N/N = 0.3$ . If increased SO<sub>2</sub> fluxes have caused a CCN change in the range  $0.1 \leq \Delta C/C \leq 0.5$  since the late nineteenth century, then  $\Delta Q_{\text{NH}}$  would lie in the range  $-0.5$  to  $-2.5 \text{ W m}^{-2}$  (or 1/0.6 times these values if averaged over only the Northern Hemisphere oceans). At the top end of this range, the perturbation is similar in magnitude to the forcing from changes in greenhouse-gas concentrations over the past few centuries (about  $2.2 \text{ W m}^{-2}$  at the top of the troposphere<sup>10</sup>).

Given the possibility of a significant SO<sub>2</sub>-related forcing in the Northern Hemisphere, it is pertinent to ask the following

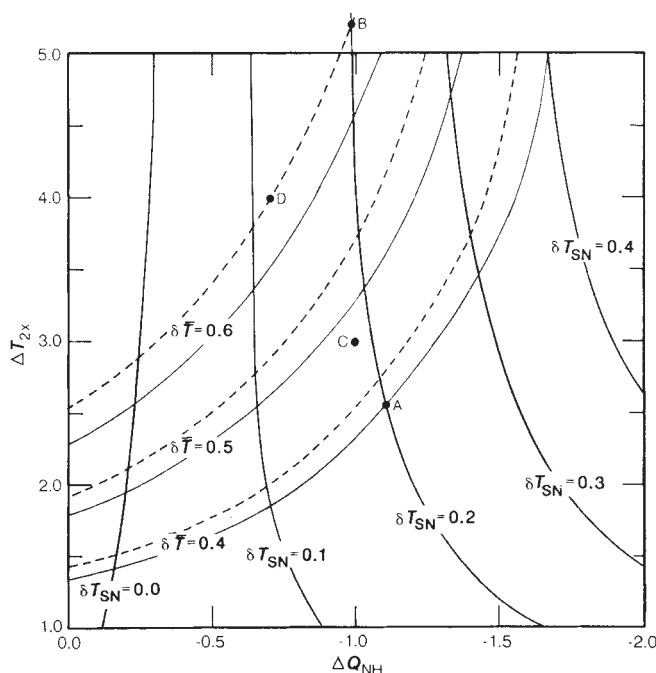


FIG. 1 The effect of climate sensitivity ( $\Delta T_{2x}$ ) and Northern Hemisphere cloud-albedo-derived forcing ( $\Delta Q_{\text{NH}}(1985)$ ) on global-mean temperature change between 1900 and 1985 ( $\delta\bar{T}$ ;  $\delta\bar{T}(1985)$  in text) and the relative warming of the hemispheres over this period ( $\delta\bar{T}_{\text{SN}}$ ;  $\delta\bar{T}_{\text{SN}}(1985)$  in text). Full lines show results for a diffusivity of  $1.0 \text{ cm}^2 \text{ s}^{-1}$ , and dashed lines are for a diffusivity of  $2.0 \text{ cm}^2 \text{ s}^{-1}$ . The effect of diffusivity on  $\delta\bar{T}_{\text{SN}}$  is negligible. Points A–D are explained in text.



questions: For a given forcing  $\Delta Q_{\text{NH}}$ , what hemispheric temperature differential would be expected? What  $\Delta Q_{\text{NH}}$  values are compatible with observed differences in changes in hemispheric-mean temperature? Could  $\text{SO}_2$ -related forcing have noticeably offset the greenhouse effect? These questions can be answered by comparing modelled and observed differences in hemispheric-mean temperature. To do this, the climate model must include at least the damping effect of oceanic thermal inertia and must differentiate the two hemispheres. It is also important to use a reasonable parameterization of inter-hemispheric heat and mass exchanges, because although hemisphere-specific forcing will produce a response that is greater in the hemisphere concerned, the response will be far from negligible in the other hemisphere because of these exchanges<sup>11</sup>. In the model used here<sup>11</sup>, this is ensured by calibrating the land-sea and hemisphere-hemisphere exchange coefficients to give the correct seasonal cycles of temperature over the land and ocean areas of each hemisphere. The adequacy of the parameterization has been further tested by simulating volcanic eruptions in both hemispheres and comparing the results with hemispheric-mean temperature observations<sup>12</sup>.

The magnitude of the model's response to a given external-forcing change depends primarily on two factors, the climate sensitivity or gain<sup>13</sup> (determined here by the equilibrium global-mean warming produced by a doubling of the  $\text{CO}_2$  concentration,  $\Delta T_{2x}$ ) and the rate of ocean mixing (controlled by the ocean's vertical diffusivity,  $\kappa$ ). For any prescribed  $\text{SO}_2$ -derived forcing history, therefore, there will be a range of possible inter-hemispheric temperature differences and global-mean temperature changes because of uncertainties in  $\Delta T_{2x}$  and  $\kappa$ .

For the forcing history, I have superimposed  $\text{SO}_2$ -derived forcing on the combined effect of changes in greenhouse-gas concentrations ( $\text{CO}_2$ ,  $\text{CH}_4$ ,  $\text{N}_2\text{O}$  and the CFCs) as given in ref. 10. Although other external forcings have been proposed<sup>14-17</sup>, greenhouse forcing is the only one that has a known history over the past century. The  $\text{SO}_2$ -derived forcing is here characterized by its 1985 value ( $\Delta Q_{\text{NH}}(1985)$ ), for which I will consider a range of values, but the response may also depend on the history of  $\Delta Q_{\text{NH}}$ . Although this is unknown, it must be related to the history of  $\text{SO}_2$  flux changes, which show a relatively slow increase over 1850–1950 and a more rapid increase since then<sup>6</sup>. Another proxy for  $\Delta Q_{\text{NH}}$  is the Greenland ice-core n.s.s.- $\text{SO}_4^{2-}$  record<sup>4,5</sup>, which shows an increase over 1895–1915, little change to around 1940 or 1955, an increase to 1970 and little change since then. Neither  $\text{SO}_2$  flux nor n.s.s.- $\text{SO}_4^{2-}$  will accurately portray changes in CCN, albedo or radiative forcing, because of possible saturation and nonlinear effects. I have investigated a number of different possible forcing histories, and all produce quantitatively similar results provided that they end at the same  $\Delta Q_{\text{NH}}(1985)$  value. The particular history whose results are described below parallels the n.s.s.- $\text{SO}_4^{2-}$  record in ref. 4 (their Fig. 2), beginning at zero in 1895, rising to a maximum in 1970 and then remaining constant. Although the relation of n.s.s.- $\text{SO}_4^{2-}$  to CCNs has been used previously by Legrand *et al.*<sup>18</sup>, it can be considered only a rough approximation, not least because the relationship is known to be nonlinear<sup>19</sup>.

The key model results are the global-mean temperature change and the change in the temperature difference between the hemispheres. All model runs begin in equilibrium in 1765. The changes described below are all relative to values in 1900 in order to facilitate comparison with observations. The global-mean temperature change is

$$\delta \bar{T}(t) = \Delta \bar{T}(t) - \Delta \bar{T}(1900) \quad (4)$$

and the change in the temperature difference between the hemispheres is

$$\delta T_{\text{SN}}(t) = (\Delta T_{\text{S}}(t) - \Delta T_{\text{N}}(t)) - (\Delta T_{\text{S}}(1900) - \Delta T_{\text{N}}(1900)) \quad (5)$$

where  $\Delta$  is the change since 1765, an overbar denotes a global average and subscripts N and S refer to the hemispheric

averages. The  $\text{SO}_2$ -related forcing was applied only over the Northern Hemisphere ocean area, with magnitude scaled from  $\Delta Q_{\text{NH}}$  by 1/0.6. Results are shown in Fig. 1. Climate sensitivity is the main controlling model parameter, although this has little effect on  $\delta T_{\text{SN}}(1985)$  for  $\Delta T_{2x} \geq 2^\circ\text{C}$ . For the chosen forcing history,  $\delta T_{\text{SN}}(1985)$  is virtually independent of diffusivity.  $\delta T_{\text{SN}}(1985)$  increases roughly linearly with increasing magnitude of  $\Delta Q_{\text{NH}}(1985)$ , and global-mean temperature change ( $\delta \bar{T}(1985)$ ) decreases markedly as the magnitude of  $\Delta Q_{\text{NH}}(1985)$  increases.

These results can be compared with observations, but there are uncertainties in such a comparison because of, on the observational side, changes in (or incomplete) data coverage, and because of, on statistical grounds, the existence of pronounced shorter-timescale (probably natural) fluctuations. Figure 2 shows  $\delta T_{\text{SN}}$  as a function of time. For the ocean areas, temperatures are based on observations of both sea surface temperatures (SSTs) and marine air temperatures (MATs) or night-time marine air temperatures (NMATs) from two different (but overlapping) data sources, the UK Meteorological Office marine data set<sup>20</sup> and the Comprehensive Ocean-Atmosphere Data Set (COADS)<sup>21,22</sup>. For the land areas, data are from Jones *et al.*<sup>23,24</sup>. Data from before about 1900 are less reliable than those after this date. The total trend over 1900–1985 is, with 95% confidence limits,  $0.03 \pm 0.13^\circ\text{C}$  (COADS MAT),  $0.14 \pm 0.14^\circ\text{C}$  (UKMO NMAT),  $0.05 \pm 0.12^\circ\text{C}$  (COADS SST) or  $0.10 \pm 0.12^\circ\text{C}$  (UKMO SST). The positive-trend values indicate cooling of the Northern Hemisphere with respect to the Southern. Autocorrelation has been accounted for in estimating the confidence limits. For comparison with the observations, Fig. 2 also shows a model result; specifically, for  $\Delta Q_{\text{NH}}(1985) = -1.0 \text{ W m}^{-2}$  with  $\kappa = 1 \text{ cm}^2 \text{ s}^{-1}$  and  $\Delta T_{2x} = 3.0^\circ\text{C}$ , which gives  $\delta \bar{T}(1985) = 0.47^\circ\text{C}$  and  $\delta T_{\text{SN}}(1985) = 0.18^\circ\text{C}$  (point C on Fig. 1).

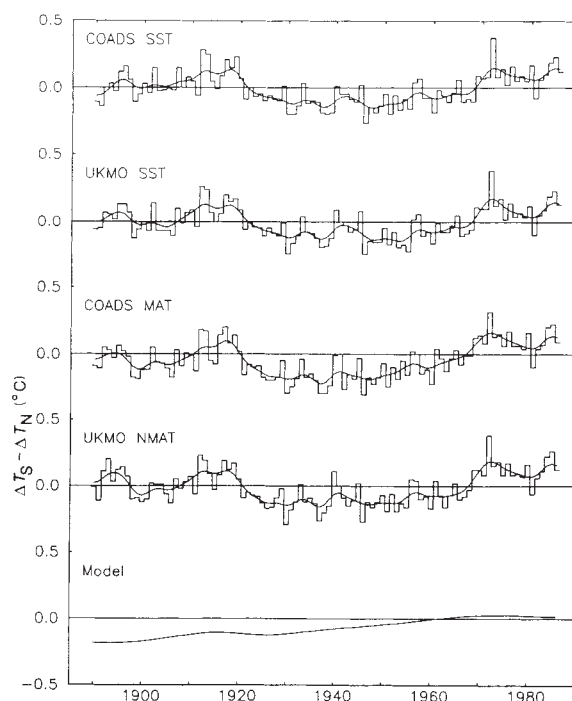


FIG. 2 Differences between changes in Northern Hemisphere temperature ( $\Delta T_{\text{N}}$ ) and Southern Hemisphere temperature ( $\Delta T_{\text{S}}$ ). A positive value indicates that the Southern Hemisphere has warmed relative to the Northern Hemisphere. Marine data from two different sources are used, as described in the text. All values are relative to a 1950–1979 reference period. The lower panel shows modelled changes in  $\delta T_{\text{SN}}$ , with the 1985 value corresponding to point C in Fig. 1. (The zero point has been shifted to give zero mean over 1950–1979.) Abbreviations as in text.

The observations indicate a  $\delta T_{SN}$  trend over 1900–1985 of around  $0.1^\circ\text{C}$ . This value is subject to considerable uncertainty, both statistical and observational, and could range anywhere between  $-0.1^\circ\text{C}$  and  $0.3^\circ\text{C}$ . For any assumed overall trend, one can estimate the required value of  $\Delta Q_{NH}(1985)$  using Fig. 1. This value depends on the assumed values of  $\Delta T_{2x}$  and  $\kappa$ , and, provided no factors other than the greenhouse effect and  $\text{SO}_2$  increases have contributed significantly to the century-timescale forcing, these quantities are constrained, in turn, by the necessity for  $\delta \bar{T}(1985)$  to be compatible with observations. From the late nineteenth century to 1985, global-mean temperature increased by about  $0.5^\circ\text{C}$ , with an uncertainty of around  $\pm 0.1^\circ\text{C}$  (ref. 22).

An example showing the constraints on  $Q_{NH}(1985)$  for given  $\delta T_{SN}(1985)$  is given in Fig. 1. If  $\delta T_{SN}(1985) = 0.2^\circ\text{C}$ ,  $0.4 \leq \delta \bar{T}(1985) \leq 0.6^\circ\text{C}$  and  $1 \leq \kappa \leq 2 \text{ cm}^2 \text{ s}^{-1}$ , then  $\Delta Q_{NH}(1985)$  must lie between the limits corresponding to points A and B in Fig. 1, specifically  $-1.0$  to  $-1.1 \text{ W m}^{-2}$ . Equation (1) implies that the required increase in average Northern Hemisphere marine CCN concentration is  $\sim 20\%$ . For  $\Delta Q_{NH}(1985) \sim -1 \text{ W m}^{-2}$ , the  $\delta \bar{T}(1985)$  limits of  $0.4$  and  $0.6^\circ\text{C}$  imply that  $\Delta T_{2x}$  must be in the range  $2.5$ – $5.2^\circ\text{C}$ . For the estimated upper limit of  $\delta T_{SN}(1985)$ ,  $0.3^\circ\text{C}$ , the hemispheric-mean  $\text{SO}_2$ -derived forcing could be as high as  $-1.4 \text{ W m}^{-2}$  (that is  $\Delta C/C \approx 30\%$ ) and a  $\Delta T_{2x}$  value close to  $5^\circ\text{C}$  would be compatible with a global-mean warming of  $0.5^\circ\text{C}$ .

The questions posed earlier can now be answered. Each  $0.1^\circ\text{C}$  increase in the twentieth-century warming of the Southern Hemisphere relative to the Northern Hemisphere corresponds to a forcing of around  $-0.5 \text{ W m}^{-2}$ , or a CCN increase of  $\sim 10\%$ . Values of up to three times these amounts would still be compatible with observations. This suggests that  $\text{SO}_2$ -derived aerosols could have had a noticeable effect on global-scale temperatures. It should be noted, however, that there are alternative explanations for differing trends in hemispheric-mean temperatures (see refs 10 and 16, for example). Furthermore, even quite large CCN changes will fail to produce a statistically significant hemispheric temperature contrast. Given the noise level and the possibility of alternative explanations, detection of a  $\text{SO}_2$ -derived CCN effect would be exceedingly difficult.

The implications of even a modest value for the  $\text{SO}_2$ -derived forcing are of considerable importance. In the absence of such forcing (and assuming no long-term forcing other than that due to greenhouse-gas concentration changes), the implied  $\Delta T_{2x}$  lies in the range  $1.3$ – $2.5^\circ\text{C}$ . These values are considerably below those given by most recent General Circulation Model (GCM) studies<sup>25</sup>. A value for  $\Delta Q_{NH}(1985)$  of as little as  $-0.7 \text{ W m}^{-2}$  (corresponding to  $\Delta C/C \approx 14\%$ ) would admit a  $\Delta T_{2x}$  value close to  $4^\circ\text{C}$ , which is more in accord with these GCM results (point D on Fig. 1); but it is doubtful that the global-mean warming since 1900 is as high as the  $0.6^\circ\text{C}$  required to give this  $\Delta T_{2x}$  value. As the magnitude of the  $\text{SO}_2$ -derived forcing increases, this increasingly offsets the greenhouse-gas forcing and allows still larger values of  $\Delta T_{2x}$ . Thus, the existence of  $\text{SO}_2$ -derived forcing may help to explain the apparent inconsistency between GCM results and observations.

An ironic twist arises if the  $\text{SO}_2$ -derived forcing is, after all, significant. The effects of  $\text{SO}_2$  associated with acidic precipitation and urban pollution are clearly detrimental, and measures to reduce emissions are being implemented widely. However, if we were successful in halting or reversing the increase in  $\text{SO}_2$  emissions, we could, as a by-product, accelerate the rate of greenhouse-gas-induced warming, so reducing one problem at the expense of increasing the rate of onset of another.  $\square$

5. Mayewski, P. A. *et al. Science* **232**, 975–977 (1986).
6. Möller, D. *Atmos. Environ.* **18**, 19–27 (1984).
7. Twomey, S. A., Piepgrass, M. & Wolfe, T. L. *Tellus* **36B**, 356–366 (1984).
8. Coakley, J. A. Jr, Bernstein, R. L. & Durkee, P. A. *Science* **237**, 1020–1022 (1987).
9. Pruppacher, H. R. & Klett, J. D. *Microphysics of Clouds and Precipitation* (Reidel, Dordrecht, 1978).
10. Wigley, T. M. L. *Geophys. Res. Lett.* **14**, 1135–1138 (1987).
11. Wigley, T. M. L. & Raper, S. C. B. *Nature* **330**, 127–131 (1987).
12. Sear, C. B., Kelly, P. M., Jones, P. D. & Goodess, C. M. *Nature* **330**, 365–367 (1987).
13. Schlesinger, M. E. *Clim. Dynam.* **1**, 35–51 (1986).
14. Hansen, J. *et al. Science* **213**, 957–966 (1981).
15. Gilliland, R. L. *Climatic Change* **4**, 111–131 (1982).
16. Gilliland, R. L. & Schneider, S. H. *Nature* **310**, 38–41 (1984).
17. Vinnikov, K. Ya. & Groisman, P. Ya. *Meteorol. i Gidrol.* **1981** (11), 30–43 (1981).
18. Legrand, M. E., Delmas, R. J. & Charlson, R. J. *Nature* **334**, 418–420 (1988).
19. Leaitch, W. R., Strapp, J. W., Isaac, G. A. & Hudson, J. G. *Tellus* **38B**, 328–344 (1986).
20. Folland, C. K., Parker, D. E. & Kates, F. E. *Nature* **310**, 670–673 (1984).
21. Woodruff, S. D., Slutsky, R. J., Jenne, R. L. & Steurer, P. M. *Bull. Am. met. Soc.* **68**, 1239–1250 (1987).
22. Jones, P. D., Wigley, T. M. L. & Wright, P. B. *Nature* **322**, 430–434 (1986).
23. Jones, P. D. *et al. J. Clim. appl. Met.* **25**, 161–179 (1986).
24. Jones, P. D., Raper, S. C. B. & Wigley, T. M. L. *J. Clim. appl. Met.* **25**, 1213–1230 (1986).
25. Schlesinger, M. E. & Mitchell, J. F. B. *Rev. Geophys.* **25**, 760–798 (1987).

ACKNOWLEDGEMENTS. Supported by the US Department of Energy, Carbon Dioxide Research Division. Some of this work was carried out while the author was a visiting scientist at the National Center for Atmospheric Research (NCAR), Boulder, Colorado, USA. NCAR is sponsored by the NSF. Help from J. A. Coakley is gratefully acknowledged.

## Mobilization of radiocaesium in pore water of lake sediments

Rob N. J. Comans<sup>\*†</sup>, Jack J. Middelburg<sup>\*</sup>,  
Jan Zonderhuis<sup>†</sup>, Joost R. W. Woittiez<sup>†</sup>,  
Gert J. De Lange<sup>\*</sup>, Henk A. Das<sup>†</sup>  
& Cornelis H. Van Der Weijden<sup>\*</sup>

<sup>\*</sup> Department of Geochemistry, Institute of Earth Sciences, University of Utrecht, PO Box 80.021, 3508 TA Utrecht, The Netherlands

<sup>†</sup> Department of Chemistry and Materials Sciences, Netherlands Energy Research Foundation (ECN), PO Box 1, 1755 ZG Petten, The Netherlands

THE deposition of large amounts of radiocaesium from nuclear weapons testing and from accidents such as Chernobyl has necessitated study of the fate of these long-lived radioisotopes in the natural environment. Caesium is known to interact strongly with micaceous clay minerals in soils and sediments<sup>1–4</sup>. Radiocaesium can therefore be removed from the water column in lake systems by settling particles and surface sediments<sup>5</sup>. This process reduces its mobility and the risk of assimilation by biota. Nevertheless, there are indications that radiocaesium may be mobilized from lacustrine anoxic sediments<sup>6</sup>. Direct evidence, however, can come only from measurements on pore water from lake sediments, and here we report such measurements. Actual *in situ* solid/liquid distribution coefficients ( $K_d$  values), determined from our data, provide convincing evidence for post-depositional mobilization of particle-bound <sup>137</sup>Cs. This is probably caused by ion exchange with  $\text{NH}_4^+$  which reaches high concentrations in anoxic pore waters. The pore-water data indicate that radiocaesium is returned to the water column and thus becomes available for uptake by aquatic organisms.

In lake systems, fallout radiocaesium is taken up by settling particles and also, in shallow parts of well mixed lakes, directly by surface sediments<sup>5</sup>. In the sediment, the decomposition of labile organic matter initiates early diagenetic reactions which may lead to enhanced pore-water concentrations of phosphate, ammonia, iron, manganese and bicarbonate. The pore-water composition of the sediment therefore changes substantially.

Sholkovitz<sup>6</sup> has suggested that <sup>137</sup>Cs, associated with particles, shows post-depositional mobility as a result of these changing redox conditions. The extent of <sup>137</sup>Cs mobility in lake sediments is still unknown. This is partly because of the lack of actual measurements of <sup>137</sup>Cs in the pore water of these freshwater environments. To the best of our knowledge, <sup>137</sup>Cs activities in natural pore water have been reported only for marine sediments

Received 10 February; accepted 24 April 1989.

1. Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. *Nature* **326**, 655–661 (1987).
2. Mészáros, E. *Atmos. Environ.* **22**, 423–424 (1988).
3. Schwartz, S. E. *Nature* **336**, 441–445 (1988).
4. Neftel, E., Beer, J., Oeschger, H., Zürcher, F. & Finkel, R. C. *Nature* **314**, 611–613 (1985).

<sup>†</sup> Present address: Department of Chemistry and Materials Sciences, Netherlands Energy Research Foundation (ECN), PO Box 1, 1755 ZG Petten, The Netherlands.



in Buzzards Bay, United States<sup>7</sup>. By measuring, for the first time,  $^{137}\text{Cs}$  in both lacustrine sediments and pore water, we were able to investigate directly the occurrence of post-depositional mobilization of  $^{137}\text{Cs}$  in these environments. This study focuses on the Ketelmeer, a shallow Dutch fresh-water lake, which forms an important sedimentation area for particulate material from the River Rhine. The residence time of water in this lake is only 2 days. The water column shows no seasonal stratification and is well mixed and oxic throughout the year.

In European surface water, pre-Chernobyl  $^{137}\text{Cs}$  activities were very low ( $\sim 1 \text{ mBq l}^{-1}$ )<sup>8</sup>. In May 1986, soon after Chernobyl, values of  $(1-2) \times 10^3 \text{ mBq l}^{-1}$  were reported for Swiss surface waters<sup>5,9</sup>. These activities decreased by an order of magnitude in the first month after the accident<sup>5,9</sup>. To measure accurately this radioisotope at such low dissolved activities, one needs to preconcentrate  $^{137}\text{Cs}$  from relatively large volumes of water (typically 10–100 l). It is not possible to collect similar quantities of pore water from sediment horizons only a few centimetres thick while maintaining *in situ* redox conditions. Because the two last conditions were necessary for this study, we preconcentrated  $^{137}\text{Cs}$  from 100-ml samples and made use of an ultra-low-background counting facility<sup>10</sup>. This procedure enabled us to measure the  $^{137}\text{Cs}$  pore-water activities with  $\sim 15\%$  standard deviation<sup>10</sup>.

The  $^{137}\text{Cs}$  profile of the sediments from the core collected on 12 November 1987 (Fig. 1a) shows two pronounced signatures: (1) a maximum at 18 cm depth related to the fallout resulting from nuclear weapons testing in the 1950s and 1960s; and (2) a near-surface enrichment related to the Chernobyl accident in April 1986. These two sources are confirmed by the  $^{134}\text{Cs}$  profile:  $^{134}\text{Cs}$  is present at relatively high activities in the top 10 cm of the core, with a  $^{134}\text{Cs}/^{137}\text{Cs}$  ratio of 0.5–0.6, which is characteris-

tic of the Chernobyl fallout<sup>11,12</sup>. Below a depth of 10 cm,  $^{134}\text{Cs}/^{137}\text{Cs}$  ratios are close to zero; the  $^{134}\text{Cs}$  activity is negligible. This indicates that radiocaesium at these depths must originate from atom-bomb tests, because  $^{134}\text{Cs}$  released in that period<sup>13,14</sup> has decayed away (for  $^{134}\text{Cs}$ , the half-life  $t_{1/2} = 2.06 \text{ yr}$ , and for  $^{137}\text{Cs}$ ,  $t_{1/2} = 30.17 \text{ yr}$ ). The near-surface enrichment pattern is typical for bioturbational mixing of an impulse source<sup>15</sup>, in this case Chernobyl. A sedimentation rate of  $0.8 \text{ cm yr}^{-1}$  is calculated on the assumption of a 1963 maximum in the fallout of  $^{137}\text{Cs}$  from atom-bomb tests. This is in excellent agreement with an estimated average sedimentation rate of  $\sim 0.8\text{--}0.9 \text{ cm yr}^{-1}$  for this part of the lake<sup>16</sup>.

Pore-water  $^{137}\text{Cs}$  activities (Fig. 1b) show no major fluctuations in the top 20 cm of the core, but there is a tendency towards depletion with depth in the 20–35-cm zone. The activity in pore water is about 1.5–2 times that of the bottom water overlying the sediment. The values shown in Fig. 1b are about an order of magnitude higher than those reported for bottom water (April 1987) in Lake Zürich, Switzerland<sup>5</sup>, and are similar to the values measured in September 1986 in Lake Lugano, Switzerland<sup>9</sup>. These observations are consistent with the  $^{137}\text{Cs}$  inventories of these lakes:  $26 \text{ kBq m}^{-2}$  (Ketelmeer)<sup>17</sup>,  $4 \text{ kBq m}^{-2}$  (Lake Zürich)<sup>5</sup> and  $20 \text{ kBq m}^{-2}$  (Lake Lugano)<sup>9</sup>.

A dissolved  $^{137}\text{Cs}$  enrichment of interstitial waters relative to bottom waters and a decrease in pore-water  $^{137}\text{Cs}$  with depth were also reported for the marine environment of Buzzards Bay<sup>7</sup>. On the basis of the  $^{137}\text{Cs}$  activity gradient between pore and bottom water, we can estimate, by means of Fick's first law<sup>15</sup>, a diffusional  $^{137}\text{Cs}$ -flux to the overlying water column of about  $0.03 \text{ Bq cm}^{-2} \text{ yr}^{-1}$ . A value of  $1.8 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  was taken for the free diffusion coefficient of  $^{137}\text{Cs}$  (ref. 18), and a value of 0.8 for the porosity of the sediment. Given a total sediment

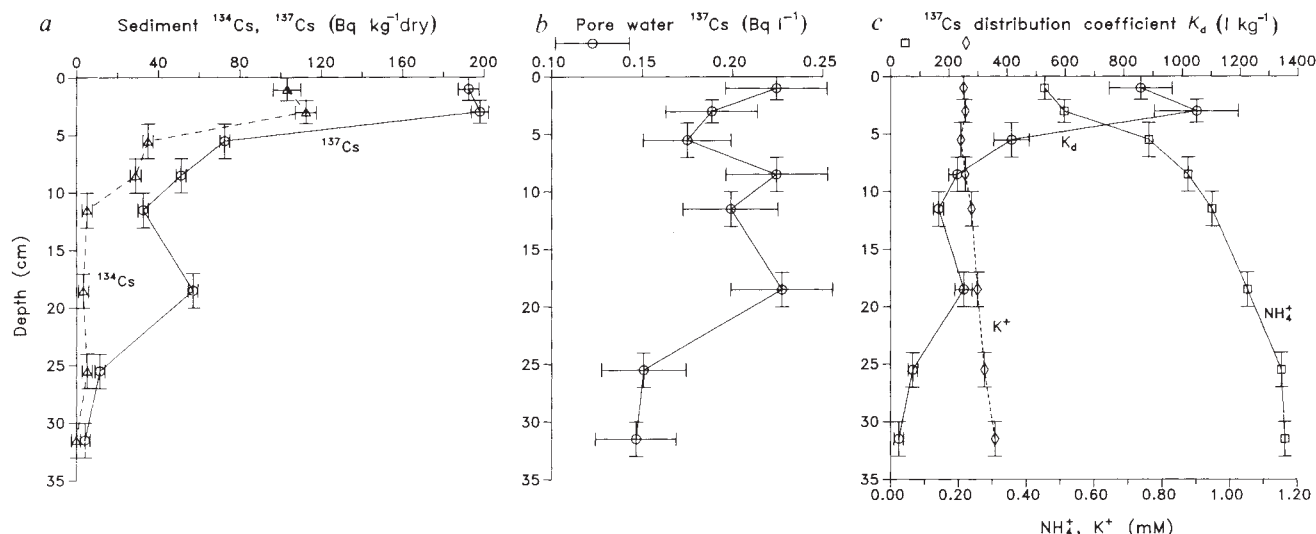


FIG. 1 Vertical profiles from the box-core taken in the fresh-water lake Ketelmeer on 12 November 1987. *a*,  $^{134}\text{Cs}$  and  $^{137}\text{Cs}$  activities in the sediment; *b*,  $^{137}\text{Cs}$  activities in pore water; *c*, *in situ*  $^{137}\text{Cs}$  distribution coefficients ( $K_d$  values) and pore-water  $\text{K}^+$  and  $\text{NH}_4^+$ . The activity of  $^{137}\text{Cs}$  and the concentrations of  $\text{K}^+$  and  $\text{NH}_4^+$  in bottom water overlying the sediment are indicated by data points above the 0-cm line.

**METHODS.** Sediments were collected with a box-corer on 12 November 1987 in Lake Ketelmeer at about 4 m water depth. Four vertical sub-cores were taken from the box-core by carefully inserting PVC tubes with a polyethylene liner. The sub-cores were sampled at identical depth intervals in a glove-bag under a nitrogen atmosphere. Pore water was extracted in a glove box close to *in situ* temperature ( $10^\circ\text{C}$ ) under a high-purity nitrogen atmosphere ( $\text{O}_2 < 0.003\%$ )<sup>23</sup>. The four sub-cores were analysed separately for nutrients, and for major and some trace elements by routine inductively coupled plasma emission spectrometry and autoanalyser techniques. No major differences were observed between identical intervals in the separate cores. This confirms the quality of our pore-water extraction and analytical techniques, and implies that the box-core showed no lateral compositional

variations. This vindicates our strategy of combining the four pore-water samples from each of the sampled depth intervals. Thus aliquots of 100 ml pore water became available for radiocaesium analysis after the addition of stable caesium as a carrier and preconcentration on ammoniummolybdophosphate (AMP)<sup>10</sup>. Low-level  $^{137}\text{Cs}$   $\gamma$ -ray counting of the preconcentrated pore-water samples was performed in an anticoincidence (AC) array consisting of an annular NaI shield around an intrinsic germanium detector. The AC facility was situated in a low-background counting room with walls of low-potassium concrete. The pore-water samples were counted for 72 h. This procedure enables  $^{137}\text{Cs}$  to be determined down to  $\sim 10 \text{ mBq}$  with  $\sim 15\%$  standard deviation and no appreciable bias<sup>10</sup>. The sediment samples were counted for 8–14 hours in aliquots of 15 g (uppermost two samples) or 30 g, with a coaxial Ge(Li) detector situated in the same low-background counting room. Both  $^{134}\text{Cs}$  and  $^{137}\text{Cs}$  activities were measured in the sediment samples. All activities were corrected for decay back to 1 May 1986. Further details of the low-level counting procedure for  $^{137}\text{Cs}$  are given in ref. 10.



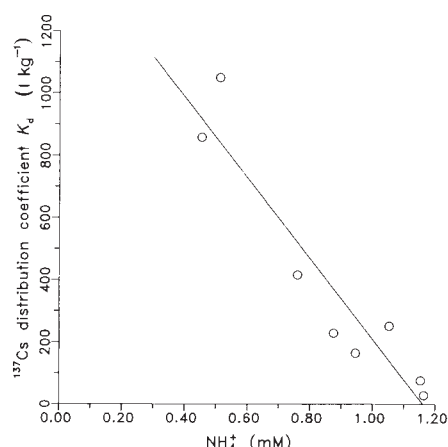


FIG. 2 Correlation of *in situ*  $^{137}\text{Cs}$  distribution coefficients between sediment and pore water ( $K_d$  values) with pore-water  $\text{NH}_4^+$  concentrations. A least-squares fit to the data gives  $K_d$  ( $\text{l kg}^{-1}$ ) =  $-1,294 [\text{NH}_4^+](\text{mM}) + 1503$  ( $r = -0.94$ ).

inventory of about  $26 \text{ kBq } ^{137}\text{Cs m}^{-2}$ , deposited in the Ketelmeer after the Chernobyl accident<sup>17</sup>,  $^{137}\text{Cs}$  from Chernobyl will be released to the water column for several decades. This first-order estimate does not take into account the role of factors such as bioturbation, advection, sediment resuspension and radioactive decay. There may also be a (smaller) downward flux of  $^{137}\text{Cs}$ , because of the slightly lower dissolved  $^{137}\text{Cs}$  activities at the bottom of the profile, as was also observed in Buzzards Bay<sup>7</sup>.

The radiocaesium data of both sediment and pore water (Fig. 1) enable us to calculate the *in situ* distribution coefficient ( $K_d$ ) of  $^{137}\text{Cs}$  at different depths in the sediment. This  $K_d$  value quantifies the partition of  $^{137}\text{Cs}$  between the sediment and its interstitial water, and is used traditionally for modelling transport and retention of this radionuclide in sediments and soils.  $K_d$ s are commonly determined in laboratory experiments and are often difficult to relate directly to natural environments<sup>19</sup>.

The profile of the calculated  $K_d$ s shows a pronounced decrease with depth (Fig. 1c). This indicates progressive mobilization of  $^{137}\text{Cs}$  resulting from decreasing association with the solids. The partition of caesium between solid and liquid phases in natural systems has been shown to be regulated by a small number of highly selective ion-exchange sites, located at the frayed edges of micaceous clay minerals<sup>4</sup>. Univalent ions with a low hydration energy and an ionic radius similar to  $\text{Cs}^+$ , such as  $\text{K}^+$ ,  $\text{NH}_4^+$  and  $\text{Rb}^+$ , are known to compete for these sites<sup>1</sup>. In natural waters, competition is expected to come mainly from  $\text{K}^+$  and  $\text{NH}_4^+$ . Of these two ions, only  $\text{NH}_4^+$  shows a distinct gradient in the Ketelmeer pore waters (Fig. 1c). Moreover, there is a significant linear relationship (correlation coefficient =  $-0.94$ ) between the  $K_d$  values and  $\text{NH}_4^+$  concentrations (Fig. 2). X-ray diffraction and chemical analysis of the sediments indicated that illite was the dominant clay mineral, with smaller contributions of montmorillonite and kaolinite. No evidence was found of major fluctuations in the amount of micaceous (illite) clay minerals with depth. We therefore conclude that the decrease in the  $^{137}\text{Cs}$  distribution coefficient with sediment depth is caused by ion exchange with pore-water  $\text{NH}_4^+$ . This effect may also explain why published  $K_d$ s for the top layer of fresh-water sediments ( $\sim 5 \times 10^3$  to  $2 \times 10^4 \text{ l kg}^{-1}$ )<sup>8,20,21</sup>, which were calculated using radiocaesium activities in the surface water, are higher than our *in situ* values.

No data are available on the  $K_d$  value of  $^{137}\text{Cs}$  for settling particles in the bottom-water region. A value of  $\sim 1,500 \text{ l kg}^{-1}$  can, however, be calculated from the  $^{137}\text{Cs}$  activity of the sediment top layer and that of the bottom water. This value may in fact be a lower limit because, in spring of 1987, higher solid-phase activities ( $300 \text{ Bq kg}^{-1}$ ) were obtained from measurements of settling particles collected at a nearby location<sup>17</sup>. Extrapolation of the regression line (Fig. 2), using the bottom-water  $\text{NH}_4^+$  concentration, gives essentially the same  $K_d$  value by independent means. If this estimate of  $1,500 \text{ l kg}^{-1}$  is valid, the  $K_d$  value of particle-bound  $^{137}\text{Cs}$  decreases substantially once settling particles have reached the sediment. This suggests that mobilization of radiocaesium may start immediately after particle deposition.

Such a rapid mobilization could be caused by the order-of-magnitude difference between the  $\text{NH}_4^+$  concentration in the bottom water ( $0.04 \text{ mM}$ ) and the pore water from the sediment top layer ( $0.45 \text{ mM}$ ).

These results agree well with observations on  $^{137}\text{Cs}$  mobilization from bottom sediments of a man-made reservoir during seasonal anoxic conditions<sup>3,22</sup>. After chemical extraction of these sediments, Evans *et al.*<sup>3</sup> concluded that ion exchange with  $\text{NH}_4^+$  was probably the main process leading to  $^{137}\text{Cs}$  release. Their results also show that  $\sim 6\%$  of the  $^{137}\text{Cs}$  can be released from the sediments at a dissolved  $\text{NH}_4^+$  concentration of  $1 \text{ mM}$ . Using our  $^{137}\text{Cs}$  data and a porosity of  $0.8$ , we calculate a  $5\%$  increase in  $^{137}\text{Cs}$  mobilization from the surface to a depth of  $30 \text{ cm}$ . As shown in Fig. 1, the  $\text{NH}_4^+$  concentration increases to  $1.2 \text{ mM}$  over the same depth. The quantitative agreement with Evans *et al.*<sup>3</sup> is excellent.

The data presented here for  $^{137}\text{Cs}$  in pore water of lacustrine sediments provide new evidence that radiocaesium released during nuclear accidents is not definitively immobilized in fresh-water sediments. Ion-exchange processes involving  $\text{NH}_4^+$ , produced during early diagenesis, may mobilize  $^{137}\text{Cs}$  progressively with depth and cause this radionuclide to be released from the sediments to the overlying waters. Chernobyl radiocaesium may thus be mobilized from the Ketelmeer sediments for several decades. Other fresh-water sediments containing radiocaesium from nuclear accidents or atom-bomb tests may be similar. This mobilization may increase the risk of radiocaesium entering the aquatic food chain. □

Received 1 February; accepted 24 April 1989.

1. Sawhney, B. L. *Clays Clay Miner.* **20**, 93–100 (1972).
2. Francis, C. W. & Brinkley, F. S. *Nature* **260**, 511–513 (1976).
3. Evans, D. W., Alberts, J. J. & Clark, R. A. *Geochim. cosmochim. Acta* **47**, 1041–1049 (1983).
4. Cremers, A., Elsen, A., De Preter, P. & Maes, A. *Nature* **335**, 247–249 (1988).
5. Santschi, P. H. *et al. Envir. Sci. Technol.* **22**, 510–516 (1988).
6. Sholkovitz, E. R. in *Chemical Processes in Lakes* (ed. Stumm, W.) 119–142 (Wiley, New York, 1985).
7. Sholkovitz, E. R. & Mann, D. R. *Geochim. cosmochim. Acta* **48**, 1107–1114 (1984).
8. Conkic, L. *et al. Wat. Res.* **22**, 241–243 (1988).
9. Santschi, P. H. *et al. in Proc. Symp. Radioaktivitätsmessungen in der Schweiz nach Tschernobyl und ihre wissenschaftliche Interpretation* 323–338 (Bundesamt für Gesundheitswesen, Bern, 1986).
10. Das, H. A. & Comans, R. N. J. *J. Radioanal. nucl. Chem.* (in the press).
11. Devell, L. *et al. Nature* **321**, 192–193 (1986).
12. Bondietti, E. A. & Brantley, J. N. *Nature* **322**, 313–314 (1986).
13. Palmer, H. E. & Perkins, R. W. *Science* **142**, 66–67 (1963).
14. Perkins, R. W. & Nielsen, J. M. *Nature* **205**, 866–867 (1965).
15. Berner, R. A. *Early Diagenesis* (Princeton University Press, 1980).
16. Winkels, H. J., Van Diem, A. & Driebergen, J. *De opbouw en kwaliteit van de waterbodem van het Ketelmeer*. Flevovericht. Rijkswaterstaat Directie Flevoland, Lelystad (in the press).
17. Coördinatie-commissie voor de metingen van radioactiviteit en xenobiotische stoffen *Rapportage aanvullend meetprogramma Tsjernobyl* VROM 80282/6-88 (1988).
18. Li, Y.-H. & Gregory, S. *Geochim. cosmochim. Acta* **38**, 703–714 (1974).
19. Honeyman, B. D. & Santschi, P. H. *Envir. Sci. Technol.* **22**, 862–871 (1988).
20. Santschi, P. H. *et al. Can. J. Fish. Aquat. Sci.* **43**, 60–77 (1986).
21. Mundscherk, H. *Deutsche Gewässerkundl. Mitt.* **29**, 4–13 (1985).
22. Alberts, J. J., Tilly, L. J. & Vigerstad, T. J. *Science* **203**, 649–651 (1979).
23. De Lange, G. J. *Geochim. cosmochim. Acta* **50**, 2543–2561 (1986).

ACKNOWLEDGEMENTS. We thank M. Haller for laboratory assistance and S. M. McNab for linguistic and stylistic advice.

# Energy budget analysis for Poás crater lake: implications for predicting volcanic activity

Geoff Brown\*, Hazel Rymer\*, John Dowden†, Phiroze Kapadia‡, David Stevenson\*, Jorge Barquero§ & Louis D. Morales||

\* Department of Earth Sciences, Open University, Milton Keynes MK7 6AA, UK

† Department of Mathematics, University of Essex, Colchester CO4 3SQ, UK

‡ Department of Physics, University of Essex, Colchester CO4 3SQ, UK

§ Department of Volcanology, Universidad Nacional de Costa Rica, Heredia, Costa Rica

|| Department of Geophysics, Universidad de Costa Rica, San Pedro, Costa Rica

A WORKING model for many active stratovolcanoes involves a magma column with a frozen cap, cooled by a meteoric-water hydrothermal system. Systems with such high latent and specific heat capacities may easily buffer internal temperatures and apparent surface activity during short-term changes in power output. The surface manifestation of volcanic hydrothermal systems takes the form of boiling mud pools, hot springs, fumaroles, and in about 20–30 cases worldwide, hot crater lakes<sup>1–7</sup>. The latter are rare because they require special conditions to exist: high water supply, confined fumarole discharge, low permeability substratum and effective sub-surface heat transport. Crater lakes at active volcanoes are in a state of dynamic equilibrium whereby annual water losses through evaporation and infiltration are balanced by additions due to, for example, rainfall and runoff. Any change in volcano power output will directly affect the internal energy and surface heat loss of the lake. Vaporization of water within the hydrothermal system, leading to enhanced steam discharge from fumaroles, can also absorb increased power output. For long-term

(months to years) power changes, we propose that crater-lake and fumarole discharge variations may well occur before significant signals on seismic and tilt networks are detected. As an illustration of these ideas, we consider here the recent activity at Poás volcano, Costa Rica.

Figure 1 depicts a model of the sub-surface geological structure of Poás volcano derived mainly from extensive gravity surveys combined with inferences from surface geology<sup>8,9</sup>. A narrow cylindrical magma feeder pipe is thought to penetrate an ancient caldera structure, ~4 km across, now filled with pyroclastic layers and some lava flows. Crater-centred micro-gravity variations<sup>10,11</sup> suggest that density changes within partially molten magma occur which are likely to be thermally induced, possibly by variations in the efficiency of the hydrothermal system. We suggest that the thermal boundary between frozen cap rock and a partially molten region below is maintained by a balance between convective heat input from the magma and heat extraction by the hydrothermal cooling system above (Fig. 1). Convection of deeply penetrating meteoric waters combined with surface evaporation is such a highly effective mechanism of heat transport that little cooling will occur during ascent (except by adiabatic decompression). Variations in fumarole and lake temperatures are probably directly related to heat flow (power output) across the boundary and, hence, to the penetration depth of the freely circulating hydrothermal system.

The present hot (40–60 °C during 1965–85 but increasing to 85 °C by late 1988), acid (pH < 0.1) crater lake has been the site of all historical activity<sup>12–14</sup>. Recent observations during advanced evaporation, with some exposure of the lake bed, show that sub-aqueous fumarolic discharge has been the source of heat and acidity. Significantly, in 1953, phreato-magmatic eruptions were preceded by the disappearance of the lake and it did not reform until 1965. Between 1973 and 1979, the lake was the site of frequent and intense phreatic eruptions. By 1980 these had ceased and fumarolic discharge was channelled through the adjacent pyroclastic cone (Fig. 1). A sharp rise in fumarole temperatures to 960 °C in 1981<sup>12,15</sup> was followed by

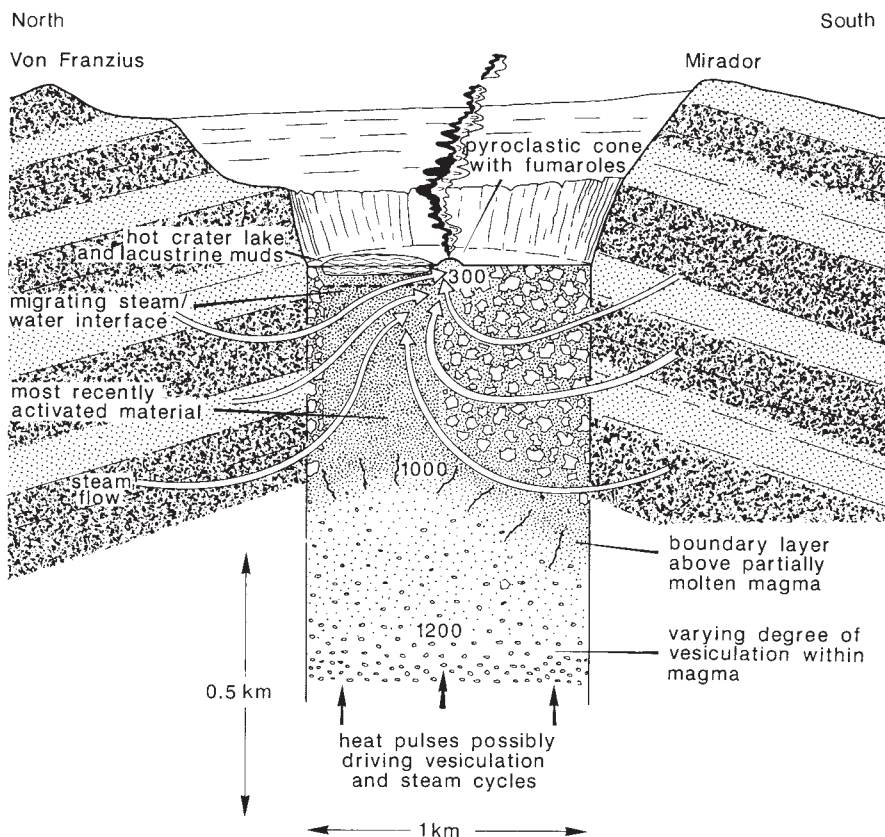


FIG. 1 Representative north-south section through the active crater at Poás as envisaged for 1985. Partially molten vesiculated magma occurs below ~500 m and is believed<sup>10,11</sup> to be undergoing ~1% variations in the exsolved gas content of crystal-free magma. Above the magma body is a zone of porous vent agglomerates within which the steam-water system is focused beneath the lake and the adjacent fumaroles. A boundary layer is envisaged where heat from a fractured, frozen cap over the partially molten magma is removed by the hydrothermal steam-water system. Numbers indicate probable temperatures in °C.

TABLE 1 Estimated power output (MW) from Poás volcano

| Type of output                                       | 1985     | Early 1988 | Late 1988    | Comments                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|------------------------------------------------------|----------|------------|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Radiation from lake surface                          | 12 ± 3   | 15 ± 5     | 10 ± 10      | Calculated from: $\alpha C(T_w^4 - T_a^4) W m^{-2}$ , where $\alpha$ (emissivity) = 0.9 for a smooth surface; $C$ (Stefan's constant) = $5.67 \times 10^{-8} W m^{-2} K^{-4}$ ; $T_w$ is the temperature of lake water and $T_a$ is the ambient temperature (10 °C for Poás).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Conductive loss to atmosphere from lake surface      | 21 ± 12  | 29 ± 15    | 18 ± 8       | Calculated from ref. 18: equal to $\rho C_p E(T_w - T_a) W m^{-2}$ , where $\rho$ (air density at 2,500 m altitude, 10 °C) = $0.919 kg m^{-3}$ ; $C_p$ (specific heat of water-saturated air at 10 °C) = $1,013 J kg^{-1} K^{-1}$ ; $E$ (the average particle flux) = 0.06 (45 °C), 0.075 (60 °C) and 0.09 (85 °C) at $2 m s^{-1}$ wind speed, down to 0.05 at $10 m s^{-1}$ ; $u$ (the friction velocity) = $0.076 m s^{-1}$ at $2 m s^{-1}$ wind speed, increasing to $0.38 m s^{-1}$ at $10 m s^{-1}$ (full discussion of $E$ and $u$ in ref. 18); $T_w$ and $T_a$ are as above. The large uncertainties quoted reflect the enormous variability in wind speed.                                                                                                                                               |
| Evaporative heat loss from lake surface              | 85 ± 50  | 175 ± 95   | 205 ± 100    | Calculated from ref. 18: $LE(q_w - q_a) W m^{-2}$ , where $L$ (latent heat of vaporization of water) = $2.39 MJ kg^{-1}$ (45 °C), $2.36 MJ kg^{-1}$ (60 °C) and $2.31 MJ kg^{-1}$ (85 °C); $q_w$ (saturation vapour concentration of water in air at lake temperature) = $0.065 kg m^{-3}$ (45 °C), $0.130 kg m^{-3}$ (60 °C) and $0.354 kg m^{-3}$ (85 °C); $q_a$ (vapour concentration in air at ambient temperature) = $0.0094 kg m^{-3}$ (10 °C) <sup>19</sup> ; $E$ and $u$ are as before and introduce similarly large uncertainties. We note that experimental data on cooling ponds <sup>20</sup> suggest that our values for evaporative heat losses from Poás crater lake may be underestimated, particularly at the higher temperatures, in which case the rise in 1988 may be even more substantial. |
| Heating runoff rainwater (60 kg s <sup>-1</sup> )    | 9 ± 2    | 13 ± 3     | 19 ± 5       | Lake temperature is important here and the quoted uncertainties reflect the variations in reported values and those we have measured.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| Fumaroles on 1953 pyroclastic cone south of the lake | 30 ± 15  | 15 ± 10    | 10 ± 10      | The largest error here is the estimation of the fumarole discharge rate, which we have only achieved crudely, and the variable fumarole temperatures (see, for example, ref. 15); we aim to quantify this aspect more rigorously in the near future.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| Summit heat flow                                     | 3 ± 1    | 3 ± 1      | 3 ± 1        | This is the radiated product of anomalously high conductive heat flow at Poás due to the hot rock at shallow depths. The estimate includes an allowance for $\leq 0.5$ MW radiation from the cone as measured using short-wavelength infrared mapping from satellite data (ref. 21 and D. A. Rothery personal communication).                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Phreatic geyser-like activity                        | —        | 15 ± 5     | intermittent | The period of regular geysering lasted from June 1987 to July 1988; there were occasional large bursts of activity (400–1,000 MW for periods up to 30 min) early in 1988. Intermittent geysering since July 1988 has not been quantified. In December 1988 parts of the lake had assumed the appearance of boiling mud pools, with much shallower steam flashing than the 40–70 m appropriate for geysers <sup>17</sup> .                                                                                                                                                                                                                                                                                                                                                                                        |
| Approximate totals                                   | 160 ± 50 | 265 ± 95   | 265 ± 100    | Statistically, the error combination gives values of uncertainty a shade larger than quoted.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |

gentle cooling (1982–86) and we measured maximum temperatures of 550 °C in 1987–88. During cooling, the magma column may develop a thickening frozen cap. Subsequently, as heat extraction by the hydrothermal system becomes less effective, the cap will be prone to fracture, thereby restoring energy extraction; fracturing in 1981 is supported by seismic data<sup>13</sup>. After a period (1982–86) during which the lake temperature remained reasonably constant, fluctuating with meteorological conditions, while the volume decreased slowly, this fracturing effect may be responsible for recent accelerated changes at Poás lake<sup>15,16</sup>.

The dominant mode of sub-surface heat transport is by rising gases which lose enthalpy, within the pipe, to its surroundings, and principally to the crater-lake zone. The main sources of heat loss are through (1) the lake surface, (2) heating cold runoff water, (3) subaerial fumarole discharges, and (4) ground surface heat flow.

(1) The radiative component is calculated<sup>5</sup> from Stefan's law (see Table 1). In 1985 we measured the surface area of the lake as  $6 \times 10^4 m^2$ ; its temperature was 45 °C, and so radiation was

12 ± 3 MW. Conduction from lake surface-water molecules to adjacent air molecules, which then convect away, depends on lake surface-wind speed, humidity and temperature (Table 1) and is estimated to be 21 ± 12 MW with large errors reflecting extremely variable weather conditions. Energy losses from lake surface evaporation are subject to similar uncertainties, but this is the dominant mode of heat loss: 85 ± 50 MW is considered a realistic estimate.

(2) The power needed to heat 60 kg s<sup>-1</sup> of rain water<sup>1</sup> from the ambient to lake temperature is 9 ± 2 MW, where the error allows for temperature and runoff variations.

(3) In 1985, a large plume of steam issued continuously from the cone (Fig. 1) and, from a crude survey of the number and estimated discharge rates of fumaroles, we deduce an output between 5 and 15 kg s<sup>-1</sup> comprising mainly high-enthalpy steam (with an average of 3 MJ kg<sup>-1</sup>). This corresponds to a power of 30 ± 15 MW with the large error resulting from temperature and volume uncertainties.

If the 1985–87 fumarole discharge rate into the lake was 80 kg s<sup>-1</sup>, as suggested by chemical-balance calculations<sup>1</sup>, the



TABLE 2 Kinetic and thermal power associated with geyser-like plumes at crater lakes, derived from the Bernoulli theorem of streamline flow coupled with the theory developed in the text

| Plume height | Plume diameter | Fluid velocity at lake level, $u$ | Exit density, $\rho_{ex}$ | Kinetic power of continuous plume | Thermal power of steam (continuous)* |
|--------------|----------------|-----------------------------------|---------------------------|-----------------------------------|--------------------------------------|
| (m)          | (m)            | (m s <sup>-1</sup> )              | (kg m <sup>-3</sup> )     | (MW)                              | (MW)                                 |
| 5            | 5              | 9.9                               | 879                       | 8.4                               | 49                                   |
| 15           | 10             | 17.2                              | 707                       | 140                               | 820                                  |
| 30           | 15             | 24.3                              | 547                       | 690                               | $4.0 \times 10^3$                    |
| 60           | 15             | 34.3                              | 376                       | $1.3 \times 10^3$                 | $7.9 \times 10^3$                    |
| 100          | 20             | 44.3                              | 266                       | $3.6 \times 10^3$                 | $2.1 \times 10^4$                    |
| 200          | 25             | 62.6                              | 153                       | $9.2 \times 10^3$                 | $5.4 \times 10^4$                    |
| 300          | 30             | 76.7                              | 108                       | $1.7 \times 10^4$                 | $1.0 \times 10^5$                    |

A geyser source depth ( $L$ ) of 50 m has been assumed here (compare with 40–70 m in ref. 17; variations of  $L$  within this range have only a small effect on calculated power).

\* Based on steam enthalpy of  $2.26 \text{ MJ kg}^{-1}$ , a minimum that serves for illustrative purposes.

minimum total power output from the lake was 180 MW, higher than but near the error limit of the  $\sim 125 \pm 50 \text{ MW}$  estimated from a combination of (1) and (2). Moreover, (1) and (2) both increased during 1985–87, giving better agreement between the surface-heat loss and chemical-balance models, an agreement which, nevertheless, may be coincidental.

(4) Summit heat flow is estimated at  $3 \pm 1 \text{ MW}$ , including 0.5 MW from the hot cone. Variations are expected to be small because of the low thermal conductivity and fairly constant saturation of the rocks.

The total power output at Poás in 1985 is therefore estimated (Table 1) at  $160 \pm 50 \text{ MW}$ . Between 1985 and early 1988 the crater-lake level dropped by 20 m with a 15% reduction of surface area; by late 1988 the level was a further 15 m lower but with an overall 65% reduction in area (to  $2 \times 10^4 \text{ m}^2$ ). Temperatures had increased to  $60^\circ \text{C}$  and  $85^\circ \text{C}$ , respectively, causing significant changes in lake surface heat losses, with particular increase in the evaporative term. Compared with 1982–85, we estimate that the number of vigorous fumaroles had decreased by about half in early 1988 and by about two-thirds towards the end of the year, probably owing to lowering of the local water table in sympathy with the lake level. There is no evidence that the amount of conductive heat flow had changed. Thus the combined total of heat loss by means of (1)–(4) in early and late 1988 (Table 1) were  $\sim 250 \pm 95 \text{ MW}$  and  $265 \pm 100 \text{ MW}$ , respectively, but with large uncertainties.

June 1987 to July 1988 was a period of geyser-like phreatic activity when, typically, plumes of muddy water and vapour were thrown up to heights of 10–20 m at intervals of 5–20 min (ref. 16), with larger directed blasts (100–200 m high plumes) on 25 January, 14 March and 9 April 1988 (our observations). The energy of a geyser is estimated by assuming that a mass  $m$  of vapour (density  $\rho_v$ ) is converted from water (density  $\rho_l$ ) and that a larger mass  $M$  of water and vapour is ejected (at density  $\rho_{ex}$ ). Energy content is underestimated, as no account of entrained sediments is made. The energy of the geyser at the zone of generation, the lake surface and the highest point, is expressed, to a first approximation, using the Bernoulli theorem (Fig. 2).

At the zone of generation, when the pressure and temperature conditions are right for geyser generation, the mass  $M$  of relatively low-density water and vapour rises adiabatically to the surface. The thermal energy  $E_{th}$  per unit mass required comprises (i) energy of expansion of vapour as pressure falls and (ii) loss of potential energy as the density contrast with surrounding water decreases. For (i),

$$E_{(i)} = \int_{z=0}^L \frac{dP}{\rho_v}$$

With the pressure at the zone of generation denoted by  $P_{gen}$  and atmospheric pressure and density denoted by  $P_{at}$  and  $\rho_{at}$  respectively, given that then

$$P_{gen} = P_{at} + \rho_l g z \text{ and } \rho_v = \frac{P_{gen} P_{at}}{P_{at}}$$

it follows by substitution that

$$E_{(i)} = \frac{P_{at} g \rho_l}{\rho_{at}} \int_{z=0}^L \frac{dz}{P_{at} + \rho_l g z}$$

Similarly,

$$E_{(ii)} = \int_{z=0}^L g \left( \frac{\rho_l}{\rho_v} - 1 \right) dz$$

and by substitution for  $\rho_v$  and  $P_{gen}$

$$E_{(ii)} = \frac{P_{at} g \rho_l}{\rho_{at}} \int_{z=0}^L \frac{dz}{P_{at} + \rho_l g z} - gL$$

The total thermal energy of the geyser is

$$E_{th} = E_{(i)} + E_{(ii)} = \frac{2P_{at}}{\rho_{at}} \ln \left[ 1 + \frac{\rho_l g L}{P_{at}} \right] - gL$$

At the lake surface, the velocity  $u$  of the vapour/water mixture is  $\sqrt{2gH}$  (Fig. 2). As the mass  $M$  ejected at the lake surface may be expressed as  $M = \rho_{ex} A t \sqrt{2gH}$ , where  $t$  is the duration of the geyser eruption and  $\rho_{ex}$  is the exit density of the ejecta, it follows that the kinetic energy of the erupted mass at the lake surface is

$$E_{k.e.} = \sqrt{2} \rho_{ex} (gH)^{3/2} A t = MgH$$

As

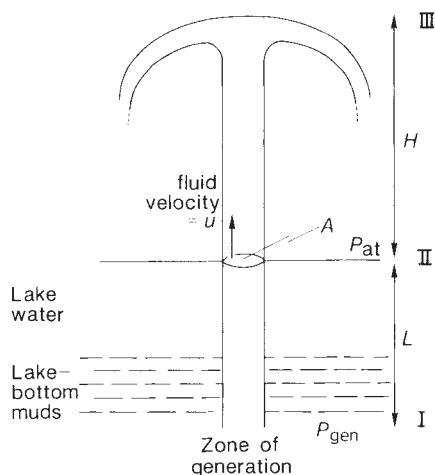
$$\frac{M}{\rho_{ex}} = \frac{M - m}{\rho_l} + \frac{m}{\rho_{at}}$$

it follows by substitution for  $M$  and the relation  $mE_{th} = MgH$  that

$$\rho_{ex} = \left\{ \frac{1}{\rho_l} + \frac{gH}{E_{th}} \left( \frac{1}{\rho_{at}} - \frac{1}{\rho_l} \right) \right\}^{-1}$$

If we assume that  $P_{at} = 10^5 \text{ N m}^{-2}$ ,  $\rho_{at} = 0.919 \text{ kg m}^{-3}$  and  $\rho_l = 10^3 \text{ kg m}^{-3}$ , then for  $L = 40\text{--}70 \text{ m}$  (ref. 17) and  $H = 15 \text{ m}$ , a typical column height in 1988, we obtain  $\rho_{ex} = 686\text{--}738 \text{ kg m}^{-3}$ , giving a continuous kinetic power of 137–147 MW for as long as the geyser operates. To simplify matters, representative calculations for a source depth  $L = 50 \text{ m}$  are given in Table 2, from which it follows that the kinetic energy of a typical 1988 geyser, 15 m high ejected from a 10-m-diameter zone (Table 2), is equivalent to a continuous power output of 140 MW, or 2.3 MW if the geyser operates for 10 s every 10 min. The additional thermal energy content of the steam required to lift the geyser is 820 MW continuously, or 13.7 MW for typical geyser operation, giving a total typical geyser power of 16 MW.

Although there are large uncertainties (Table 1 totals) we conclude that the power output of the Poás summit has probably been increasing over recent years, and this deduction is supported by recent larger, short-duration phreatic eruptions with an output of 400–1,000 MW (100–200-m plumes; see Table 2). Enhanced evaporation from the crater lake is almost certainly



Bernoulli Theorem gives :  
 $(P_{at}/\rho) + 1/2 u^2 \text{ (at II)} = (P_{at}/\rho) + gH \text{ (at III)}$

Geyser velocity at lake level (II),  $u = \sqrt{2gH}$

FIG. 2 Schematic section of Poás geyser showing parameters used in power calculations:  $H$ , height reached by geysers;  $L$ , depth of geyser generation;  $A$ , cross-sectional area of lake surface through which geyser operates;  $P_{at}$ , atmospheric pressure;  $P_{gen}$ , pressure at geyser source;  $g$ , acceleration due to gravity;  $\rho$ , density of steam/water mixture (further details in text).

because temperature increases have run ahead of surface-area reductions (Table 1) and this evaporation process more than accounts for the reduced lake volume. The return to geyser-like activity at Poás in 1987 was probably a function of both the increased thermal output and the reduction of pressure<sup>17</sup>, due to water loss, on the zone of geyser generation. Finally, while steam-water buffering of apparent surface activity at crater-lake volcanoes is a highly efficient means of maintaining unstable equilibrium during periods of short-term power-output fluctuations, recent changes at Poás may have exceeded the capacity of this particular buffer system to cool the deep magma body and so maintain its steady-state position. The implications are that maintained power output and/or low water supply could culminate in a dramatic change of activity, possibly with devastating results. □

*Note added in proof:* After continued evaporation through the dry season, Poás lake disappeared in late April 1989 accompanied by several days of continuous phreatic geysering. A dry steam/‘ash’ plume at  $\geq 450^\circ\text{C}$  (radiometer measurement) was erupted to 200 m height on 25 April; from 30 April to early May a continuous plume reached 2 km in height with fallout over 200 km<sup>2</sup>.

Received 25 October 1988; accepted 24 April 1989.

1. Brantley, S. L., Borgia, A., Rowe, G., Fernandez, J. F. & Reynolds, J. F. *Nature* **330**, 470–472 (1987).
2. Hurst, A. W. in *Volcanic Hazards Assessment in New Zealand* Vol. 10 (eds Gregory, J. G. & Watters, W. A.) 23–34 (New Zealand Geol. Surv., Lower Hutt, 1986).
3. Hurst, A. W. & Dibble, R. R. *J. Volcan. geotherm. Res.* **9**, 215–236 (1981).
4. Shepherd, J. B. & Sigurdsson, H. *J. Volcan. geotherm. Res.* **13**, 119–130 (1982).
5. Shepherd, J. B. & Sigurdsson, H. *Nature* **271**, 344–345 (1978).
6. Casadevall, T. J. et al. *J. Volcan. geotherm. Res.* **23**, 169–191 (1984).
7. Thorpe, R. S., Brown, G. C., Rymer, H., Barritt, S. & Randal, M. *Earthquake Info. Bull.* **17**(2), 44–49 (1985).
8. Rymer, H. thesis, Open Univ. (1985).
9. Brown, G. C., Rymer, H. & Thorpe, R. S. *Earth planet. Sci. Lett.* **82**, 323–334 (1987).
10. Rymer, H. & Brown, G. C. *Nature* **311**, 243–245 (1984).
11. Rymer, H. & Brown, G. C. *Bull. volcan.* **49**, 389–398 (1987).
12. Casertano, L., Borgia, A. & Cigolini, C. *Bol. Volc.* **14**, 95–97 (1984).
13. Casertano, L. et al. *Geofis. Int.* **24**(2), 315–332 (1985).
14. Prosser, J. T. & Carr, M. J. *J. Volcan. geotherm. Res.* **33**, 131–146 (1987).
15. Barquero, J. in *Historical Unrest at Large Calderas of the World* Vol. 2 (eds Newhall, C. G. & Dzurisin, D.) 878 (USGS Bulletin 1855, Washington DC, 1988).
16. *Scientific Event Alert Network Bull.* **13**(6), 13–14, **13**(9), 8–10 (1988).
17. Casertano, L. et al. *Bull. volcan.* **49**, 588–598 (1987).
18. Weisman, R. N. & Brutsaert, W. *Wat. Resour. Res.* **9**, 1242–1257 (1973).
19. Mayhew, Y. R. & Rogers, G. F. C. *Thermodynamic and Transport Properties of Fluids*, 20 (Blackwell, Oxford, 1973).

20. Ryan, P. J., Harleman, D. R. F. & Stolzenbach, K. D. *Wat. Resour. Res.* **10**, 930–938 (1974).
21. Rothery, D. A., Francis, P. W. & Wood, C. A. *Proc. 6th Thematic Conf. on Remote Sensing, Houston, Texas* **1**, 283–291 (1988).

ACKNOWLEDGEMENTS. We thank Earthwatch (The Centre for Field Research, Watertown, Massachusetts), the UK Natural Environmental Research Council and the Open University for financial support. We also appreciate reviews by S. L. Brantley, D. A. Rothery and R. S. Thorpe.

## Mantle stratification and long-term polar wander

Roberto Sabadini\* & David A. Yuen†

\* Dipartimento di Fisica, Settore Geofisica, Università di Bologna, 40127 Bologna, Italy

† Department of Geology and Geophysics and Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455, USA

THERE is accumulating evidence<sup>1,2</sup> that the average location of the Earth's magnetic pole has moved relative to the deep mantle over the past 200 Myr. This phenomenon of ‘true polar wander’ may be caused by the advance and retreat of ice sheets<sup>3</sup> or by mass redistribution in the Earth's interior due to changes in the pattern of mantle convection<sup>4,5</sup>. New analyses<sup>1,2,6</sup> of polar-wander data show a significant shift of the pole in the late Cretaceous, but this period is thought to have been too warm for glaciation to have occurred. Thus, most of the true polar wander must be due to mass movements in the mantle. Here we show, by analysing the appropriate equations for polar wander, that both viscosity and chemical stratification in the mantle are important in determining the rate of polar wander. The dynamical effects of chemical stratification and high viscosity ( $>10^{23}$  poise) in the lower mantle are to decrease polar-wander speed to a level consistent with the averaged velocities inferred from palaeomagnetic data. Whole-mantle convection models, without non-adiabatic density jumps at 670 km depth and with viscosities less than  $10^{23}$  P, would produce wandering rates in excess of  $5^\circ$  per Myr.

In the past decade, estimates<sup>7,8</sup> of the magnitude of long-term true polar wander (TPW) have been decreasing. But recent new analyses<sup>1,6,9</sup> of palaeomagnetic data show that a relatively rapid polar shift of  $10$ – $15^\circ$  took place between 70 and 100 Myr ago. These represent minimum rates of polar wander because of the poor age resolution of the palaeomagnetic data. Such a shift would have been comparable with continental drift rates. Until now, most attempts<sup>10</sup> to explain the causes of TPW by internal mass anomalies have been based solely on evaluation of the changes in the moment of inertia for point masses inside a rigid Earth. The equations of motions for polar wander should be used to study properly the influences of mass heterogeneities on TPW.

We present a model that provides a physical description of TPW induced by internal mass anomalies. Much of the development can be found in previous work<sup>3,11</sup> on external forcing due to glaciation processes, which are responsible for the present-day TPW from viscoelastic mantle rheology<sup>3,11</sup>.

We use the eulerian equations of motion to analyse rotational motions of the Earth arising from internal forcing. We introduce the dimensionless variable  $m_i = \omega_i/\Omega$  where the  $m_i$ s are the direction cosines of the rotation axis in the body-fixed coordinate system and  $\Omega$  is the diurnal angular velocity. The effect of the centrifugal deformation resulting from changes in the rotation rate, and the effect of viscous flow in the mantle as a result of mass heterogeneities inside, contribute dynamically to changes in the moment-of-inertia tensor.

A set of nonlinear, ordinary differential equations (ODEs) for  $m_i$  valid for describing large polar wander have been derived<sup>3</sup>. A comparison<sup>3</sup> between the exact nonlinear ODEs and the linearized solutions showed that they were within 10% of one another up to a polar drift of  $15^\circ$ . Thus, to quantify TPW

in the late Cretaceous, the linearized Liouville equations can be used.

A second approximation scheme<sup>12,13</sup> less restrictive than complete linearization, assumes that the Earth behaves as a fluid insofar as long-term rotation is concerned. The ODEs for  $m_i$  for viscous rheology are also nonlinear<sup>3,13</sup>. But for modest amounts of polar wander,  $\leq 20^\circ$ , one may use the linearized version<sup>3</sup> of polar wander for a viscous mantle. For large time-scales the linear equation takes the form

$$\frac{d}{dt} \mathbf{m} \approx \frac{A_1}{C-A} (\Delta I_{zx} + i \Delta I_{yz}) \quad (1)$$

where  $\mathbf{m} = (m_x, m_y)^T$  are the direction cosines of the rotation axis relative to reference axes using the Conventional International Origin<sup>3</sup>, and  $C$  and  $A$  are the polar and equatorial moments of inertia, respectively, for the Earth;  $A_1$  is defined below.

The  $\Delta I_{ij}$ s represent changes in the elements of the moment-of-inertia tensor caused by (1) top and bottom boundary distortions; (2) any internal boundary deflections, which are produced by viscous flow due to the mass anomalies; and (3) the interior mass heterogeneities themselves. Postglacial rebound only influences  $\Delta I_{ij}$  for the first two contributions.

The Earth undergoes changes in rotational speed because of modification of the moment-of-inertia tensor. The term  $A_1$  controls this rotational readjustment of a viscously stratified Earth over long timescales, exceeding  $\sim 10^6$  yr. It is given by

$$A_1 = \frac{-i\sigma \prod_{j=1}^M s_j}{\prod_{j=1}^M a_j} \quad (2)$$

where  $s_j$  and  $a_j$  are respectively the inverse relaxation times for isostatic and rotational readjustments,  $M$  is the number of modes associated with the Earth model and  $\sigma$  denotes the frequency of the Chandler wobble. The eigenspectra involving  $s_j$  and  $a_j$  are obtained in this rotational problem only for the second harmonic. We note that  $A_1$  can be extracted from the analytical solution associated with the Fermi-Gougel<sup>12</sup> problem on Earth rotation.

Inspection of equation (1) shows that the polar speed depends on the product of the perturbed moment of inertia and  $A_1$ , the drift rate resulting from the long-term rotational readjustment. Past arguments<sup>1,10,6</sup> on the magnitude of polar wander have focused mainly on the perturbations to the moment of inertia. Little attention has been paid to the role of the wander rate  $A_1$  in long-term rotational dynamics.

The term  $A_1$  depends sensitively on both the viscosity and the chemical stratification in the mantle. The model<sup>14</sup> consists

of four layers: an elastic lithosphere of 120 km thickness, two viscoelastic layers separated at 670 km depth in the mantle and an inviscid core. In this model there is no viscous coupling between the core and mantle, that is, the mantle moves relative to the core in response to mass heterogeneities. Here we are dealing with the Maxwell rheology which yields fluid-like properties over long timescales. The nature of the 670-km discontinuity is important in determining the exact boundary condition to be applied on the normal stress. For a compositional boundary this discontinuity must behave nonadiabatically. In this case the appropriate boundary condition for the normal stress at 670 km would be a jump condition with an amplitude of  $\Delta\rho gU$ , where  $\Delta\rho$  is the density difference between the lower and upper mantle,  $g$  is the acceleration due to gravity and  $U$  is the vertical displacement. The behaviour of a phase boundary is more complicated, but it has been shown<sup>15</sup> that in the presence of vigorous convection the 670-km discontinuity would behave adiabatically for a phase change. In this case the boundary condition would involve continuity. However, in the case of layered convection induced by strongly endothermic phase transitions<sup>16</sup>, the appropriate boundary condition for normal stress may be of nonadiabatic character.

In Fig. 1 we compare the  $A_1$  values for chemical (non-adiabatic) boundaries and for phase transitions (adiabatic) as a function of the lower-mantle viscosity  $\nu_{lm}$ . The value of the upper-mantle viscosity  $\nu_1$  is fixed at  $10^{22}$  P. For  $\nu_{lm}$  less than  $10^{23}$  P, polar speeds, as reflected by  $A_1$ , are much greater for a phase transition (dashed curve) than for a compositional boundary (solid curve). This difference diminishes at  $\sim 10^{24}$  P, then grows again with larger lower-mantle viscosities. The polar speed also decreases sharply with a stiffer lower mantle. We note, that for  $\nu_1/\nu_{lm} = 1$ ,  $A_1$  would decrease by another factor of five<sup>17</sup>, if the 400-km discontinuity is treated as a chemical boundary<sup>18</sup> instead of a phase transition. Results for  $A_1$  do not depend on lithospheric thickness between 50 and 200-km. Thus, polar wander would be slowest for an Earth with multiple chemical boundaries.

For  $\Delta I_{ij} = 10^{32}$  kg m<sup>2</sup>, comparable to the uncompensated values resulting from Pleistocene deglaciation<sup>11</sup> (which is associated with the initial stages of melting), and also comparable to the results from dynamic slab models<sup>19</sup> of the present-day configuration, very large polar-drift rates of  $30^\circ$  Myr<sup>-1</sup> and  $6^\circ$  Myr<sup>-1</sup> are attained using linear theory for an isoviscous mantle of  $10^{22}$  P. Only with a viscosity contrast of  $\nu_1/\nu_2 = 1/30$  are rates of  $1^\circ$  Myr<sup>-1</sup> attainable for chemically layered models.

To understand better the basic differences in  $A_1$  from these two types of boundary conditions, Fig. 2 shows the radial ( $u_r$ ) and tangential ( $u_\theta$ ) displacements for tidal potential forcing at

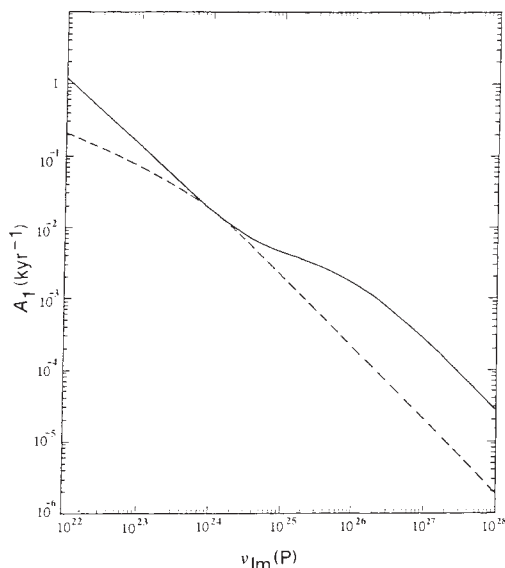


FIG. 1 The rotational excitation factor  $A_1$  as a function of lower-mantle viscosity  $\nu_{lm}$ . The upper-mantle viscosity is kept constant at  $10^{22}$  P. Dashed curve denotes phase-transition models in which there is no internal buoyancy at the 670-km discontinuity. Solid curve represents models with a non-adiabatic density jump of 9% at 670 km depth. For off-diagonal elements of the moment-of-inertia tensor  $= 10^{32}$  kg m<sup>2</sup>, a polar velocity of  $3^\circ$  Myr<sup>-1</sup> would result, when  $A_1 = 0.1$  is used in equation (1).



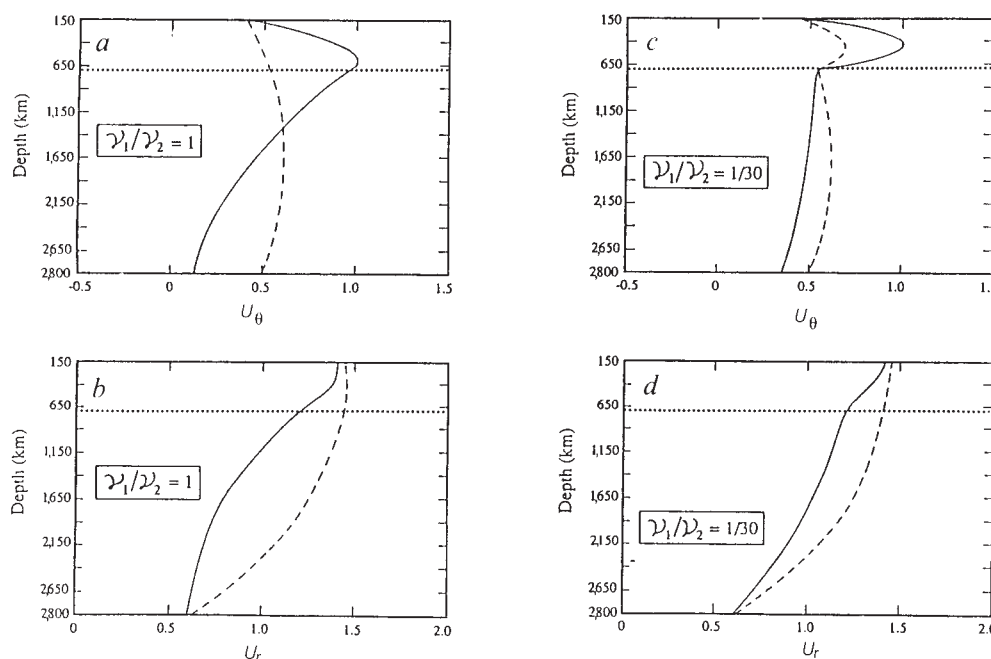


FIG. 2 Displacement fields (tangential,  $u_\theta$ , and radial,  $u_r$ , in the fluid limit) due to tidal forcing for the second harmonic. A static tidal load has been imposed as part of the surface boundary conditions, which is manifested in the gradient of the perturbed gravitational potential. The shear and normal stresses vanish at the surface. The displacement fields have been non-dimensionalized by  $\mu_1/\rho_1 d$ , where  $\mu_1$  and  $\rho_1$  are respectively the averaged shear modulus and density of the upper mantle. The viscosities of the lower and upper mantles,  $\nu_2$  and  $\nu_1$ , are  $10^{22}$  P in *a* and *b*. In *c* and *d* the lower- and upper-mantle viscosities  $\nu_2$  and  $\nu_1$  are  $3 \times 10^{23}$  and  $10^{22}$  P respectively. Horizontal dotted line denotes the 670-km discontinuity.

the surface. This forcing is related to rotational readjustment from MacCullagh's formula<sup>12</sup>. The fields plotted in Fig. 2*a, b* for an isoviscous mantle at  $10^{22}$  P reveal a much larger excitation for the adiabatic boundary condition (dashed curves), particularly in the deep mantle. The tangential displacement for the chemically stratified model (solid curve) peaks near the discontinuity because of the internal buoyancy force there. In Fig. 2*c* and *d* we show the effects on the tidal displacements from an increase of lower-mantle viscosity to  $3 \times 10^{23}$  P. We observe that the differences in the displacement fields between the two boundary conditions become smaller. This then accounts for the smaller differences in  $A_1$  (see Fig. 1), as the lower-mantle viscosity is increased.

The tangential displacement in the upper mantle for the chemically stratified model is amplified both by viscosity stratification and by internal buoyancy from the interface deflection. From Fig. 2 we can discern that the radial derivatives of the tangential displacements are greatly influenced by the rheological and chemical stratification of the mantle. These results strongly hint at the potential importance of rheology and mantle chemistry in the generation of membrane stresses<sup>20,21</sup> associated with polar wander, which hitherto have been based solely on geometrical arguments drawn from thin-shell theory<sup>20</sup>.

Besides the differences in  $A_1$ , there are also significant differences in the dynamically induced  $\Delta I_{ij}$  from internal mass anomalies in the two models with chemical and phase-transition boundaries. This can be shown by making use of the geoid Green's functions<sup>22,23</sup>  $\Delta N$  for internal loading, derived for viscous Earth models. Using MacCullagh's formula from potential theory<sup>24</sup>, we can relate changes of the second-order ( $l=2$ ) components for  $\Delta N$  to  $\Delta I_{ij}$ .

For a point mass located at depth  $d$  we can determine the changes in the moment of inertia for a viscous mantle by using the amplitude of the geoid Green's function  $\Delta N$  for a radius  $r=d$ . From calculated geoid Green's functions<sup>22,23</sup> one finds that these functions have much larger amplitudes in the phase-transition model than in chemically layered models.

At depths between 300 and 500 km, where there may be additional mass anomalies resulting from thermally induced phase transitions<sup>25</sup>, the geoid Green's functions are 3–5 times larger for phase-transition (whole-mantle) models. Thus, changes in the moment of inertia are greater in the phase-transition models. Both factors,  $A_1$  and  $\Delta I_{ij}$ , act together to make polar-wander speeds in whole-mantle models even more

unacceptably large for viscosities  $\leq 10^{23}$  P. In the case of thermally induced phase changes<sup>18</sup>, one would require lower-mantle viscosities of at least  $10^{24}$  P to decelerate TPW velocities to minimum levels<sup>6</sup> of between 0.5 to  $1^\circ$  Myr<sup>-1</sup>. On the other hand, viscosities in the lower mantle, slightly exceeding  $10^{23}$  P, are needed to achieve the averaged polar-drift rates inferred for the late Cretaceous in the case of chemically stratified models.

Our results support the case for a stiffer lower mantle, with a viscosity exceeding  $10^{23}$  P, which have been inferred both from geoid anomaly studies<sup>19</sup> and from postglacial rebound studies with transient rheology<sup>26</sup>. We emphasize that long-term TPW speeds can put strong constraints on both the chemical stratification and the viscosity distribution in the mantle and that there are trade-offs between the style of mantle chemical stratification and the magnitude of lower-mantle viscosity. Extremely fast rates for lower-mantle viscosities  $\leq 10^{23}$  P would allow the pole to wander very fast. We therefore propose that improved age resolution of palaeomagnetic data in the past 100 Myr will be useful in answering questions regarding the nature of mantle convection. □

Received 11 November 1988; accepted 18 April 1989.

- Gordon, R. G. & Livermore, R. A. *Geophys. J. R. astr. Soc.* **91**, 1049–1057 (1987).
- Sager, W. W. & Bleil, U. *Nature* **326**, 488–490 (1987).
- Sabadini, R. & Peltier, W. R. *Geophys. J. R. astr. Soc.* **66**, 533–578 (1981).
- Keondzhyan, V. P. & Monin, A. S. *Izv. Phys. Solid Earth* **13**, 760–772 (1977).
- Goldreich, P. & Toomre, A. *J. geophys. Res.* **74**, 2555–2567 (1968).
- Gordon, R. G. *Rev. Earth planet. Sci.* **15**, 567–593 (1987).
- Jurdy, D. M. & van der Voo, R. *Science* **187**, 1193–1196 (1975).
- Jurdy, D. M. *Tectonophysics* **74**, 1–16 (1981).
- Gordon, R. G. *Geophys. Res. Lett.* **10**, 709–712 (1983).
- Jurdy, D. M. *J. geophys. Res.* **88**, 6395–6402 (1983).
- Sabadini, R., Yuen, D. A. & Boschi, E. *J. geophys. Res.* **89**, 7609–7620 (1984).
- Munk, W. H. & MacDonald, G. J. F. *The Rotation of the Earth* (Cambridge University Press, 1960).
- Burgers, J. M. *Proc. K. ned. Akad. Wet.* **58**, 219–237, 1955.
- Yuen, D. A., Sabadini, R. & Boschi, E. *J. geophys. Res.* **87**, 10 745–10 762 (1982).
- Christensen, U. R. *J. geophys. Res.* **90**, 11 312–11 318 (1985).
- Christensen, U. R. & Yuen, D. A. *J. geophys. Res.* **90**, 10 291–10 300, (1985).
- Wu, P. & Peltier, W. R. *Geophys. J. R. astr. Soc.* **76**, 753–791 (1984).
- Anderson, D. L. & Bass, J. D. *Nature* **320**, 321–328 (1986).
- Hager, B. H. *J. geophys. Res.* **89**, 6003–6015 (1984).
- Turcotte, D. L. *Geophys. J. R. astr. Soc.* **36**, 33–42 (1974).
- Dickman, S. R. & Williams, D. R. *Geophys. Res. Lett.* **8**, 199–202 (1981).
- Richards, M. A. & Hager, B. H. *J. geophys. Res.* **89**, 5987–6002 (1984).
- Ricard, Y., Fleitout, L. M. & Froidevaux, C. *Ann. Geophys.* **2**, 267–286 (1984).
- Heiskanen, W. A. & Moritz, H. *Physical Geology* Ch. 2 (Freeman, San Francisco, 1967).
- Anderson, D. L. *J. geophys. Res.* **92**, 13 968–13 980 (1987).
- Yuen, D. A., Sabadini, R. C. A., Gasperini, P. & Boschi, E. *J. geophys. Res.* **91**, 11420–11438 (1986).

ACKNOWLEDGEMENTS. We thank Drs B. Hager and S. Honda for discussions and Dr S. Dickman for a review. This work was supported by the Italian Space Agency of the National Council of Research, NASA and the NSF. We thank A. Boyd for help in preparing this manuscript.

# Proterozoic microfossils from manganese orebody, India

P. C. Bandopadhyay\*

Geological Survey of India, Calcutta, India

THE presence of carbonaceous matter, and stromatolitic structures in manganese deposits of Proterozoic (>600 Myr) age<sup>1-3</sup>, provides indirect evidence for the activity of living organisms during their formation. Here, I report the presence of microfossils from the Proterozoic manganese orebody of the Penganga Group, India. These microfossils occur commonly in chert and are present within manganese oxide sediments. The microfossils probably have affinities with the Cyanobacteria and may have been planktonic; both coccoid and filamentous forms are present. The presence of these microfossils suggests a possible genetic relationship between microbiota and manganese deposition and provides further information about Proterozoic ecology.

The Penganga Group, one of several middle to late Proterozoic sedimentary sequences recognized within the intracratonic Godavari rift basin<sup>4-6</sup>, represents a marine sequence well exposed around Adilabad (Fig. 1). This sequence contains a well bedded manganese orebody conformably enclosed in a micritic limestone formation overlain by a shale formation and underlain by a near-shore crossbedded sandstone<sup>7</sup>. The presence of a very thick and persistently bedded limestone-shale sequence containing debris flow deposits and devoid of any shallow water feature suggests that deposition occurred in relatively deep-water<sup>8</sup>. The limestone is locally interbedded with glauconitic sandstone. Authigenic glauconite from this sandstone has yielded ages of  $775 \pm 30$  and  $790 \pm 30$  Myr<sup>8</sup>. The 1-10-mm thick distinct interbanding of manganese oxide and chert (Fig. 2) on a regional scale that constitutes the orebody, provides further evidence that deposition took place in a calm and quiet environment in the deeper part of a stable shelf. The orebody exposed in quarry sections is up to 0.7-m thick and is delimited above and below by unweathered limestone<sup>9</sup>. The ores are predominantly microcrystalline in structure and consist of oxide minerals<sup>10</sup>. The chert is comprised of microcrystalline quartz (5-50- $\mu$ m in size) and is made up of a fine-grained mosaic with locally recrystallized and replaced fabric. A detailed study of the mineralogy and texture of the ores, chert and host limestone confirms that the orebody was formed as sediments and modified only under diagenetic conditions, escaping further metamorphism<sup>7,10,11</sup>.

I have identified a small assemblage of coccoid and filamentous microfossils exhibiting possible affinities with the Cyanobacteria. These microfossils are structurally and organically preserved by permineralization of microcrystalline silica, similar to those of Gunflint<sup>12</sup> and the Bitter Spring microflora<sup>13</sup>. They are totally absent in recrystallized or fracture-filling fabric. These structures, also replaced or encrusted by late diagenetic manganese oxide, are comparable to the mineralized microfossils reported from Beck Spring<sup>14</sup> and McArthur deposits<sup>15</sup>.

The unicell-like bodies are amber to brown spheroids (diameter 4-24  $\mu$ m, average 13  $\mu$ m, mode 10  $\mu$ m) with smooth organic walls and fine to coarse reticulate surface textures attributable to degradation and diagenesis<sup>16,17</sup> (Fig. 3a, a', b). They exhibit budding characteristics (Fig. 3b, c) similar to those of *Huronispora* spp., suggesting a possible mechanism of multiplication<sup>18</sup>. The cells are usually well preserved showing only a little shrinkage, while the extra-cellular sheaths are not visible (Fig. 3c). This possibly suggests an uncommon preservation-degradation mode<sup>19</sup>. The cells occur as loose aggregates of about 50 or as unicells and lack a true colonial habit. Neither do they

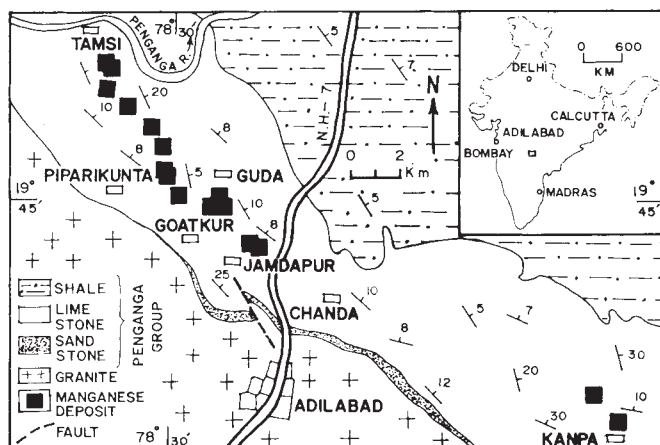


FIG. 1 Geological map of the area around Adilabad showing the distribution of the manganese orebody.

seem to have taken part in *in situ* cell division. Taken together these facts suggest that the cells were planktonic<sup>20</sup>.

The structures differ markedly in colour and morphology from synthetic quartz spherulite<sup>21</sup>. Moreover, their size distribution (Fig. 3d) is similar to those of many extant and fossil prokaryotic microbial groups<sup>22,23</sup>. The moderately large size range suggests the absence of any sampling artefact<sup>24</sup>.

These observations provide cogent evidence for the biogenic origin and possible cyanobacterial affinity of these spheroids. But, because of the difficulties of ascribing taxonomic relationships purely on the basis of morphology<sup>25</sup>, their affinity to groups other than the Cyanobacteria cannot as yet be ruled out.

Like the spheroids, the filamentous structures tend to occur as single entities rather than interwoven mats. They are straight or slightly curved, tubular to partially flattened cylindrical structures (Fig. 3e). They are apparently aseptate and unbranched, and exhibit distinct yellowish brown remnants of what could possibly have been terminal cells. They are pigmented by organic matter, light to dark brown in colour, exhibit a fine to coarse granular surface texture, and a distinct lateral wall of consistent thickness (1-2  $\mu$ m) to an intermittent or poorly defined boundary. These textural variations may reflect differences in post mortem degradation<sup>17,26-28</sup>. I have also observed truncation or indentation along the sheath margin, possibly due to growth of the mineral chalcedony, comparable to diagenetically deformed sheaths<sup>26,29</sup>. The morphology of the structures, their mode of occurrence and their size distribution (diameter 14-40  $\mu$ m, average 27  $\mu$ m; Fig. 3f) suggest that these tubes may represent discarded sheaths of broad planktonic filamentous cyanobac-

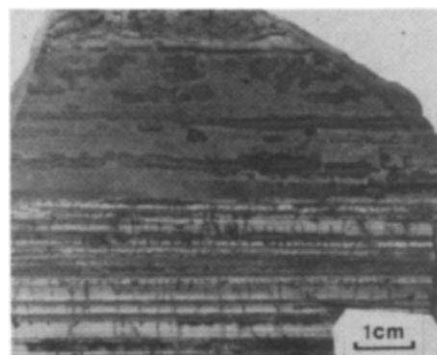


FIG. 2 Interbedded manganese oxide-chert (white, bluish grey, red and black) constituting the orebody. The top chert band contains numerous small brown, lenticular to irregular dots and is nonlaminated, whereas the basal chert bands are simply laminated. Microfossils are more common in the top band. Note the presence of syneresis cracks within chert bands filled with late diagenetic Mn oxide.

\* Address for correspondence: 40 Baje Sibpur Road, Howrah 711102, West Bengal, India

teria similar to those reported from Late Precambrian sediments of Soviet Union<sup>20</sup>.

An important observation is the occurrence of microbiota within white to grey primary manganese oxide sediments. These are loosely aggregated elliptical to rod-shaped, cell-like units which are unornamented and dark in colour (Fig. 4a). The cells are between 8 and 30  $\mu\text{m}$  long and between 3 and 12  $\mu\text{m}$  wide (Fig. 4d). It is not clear whether these structures have affinities with the Cyanobacteria, although their relatively large size would

tend to favour such an interpretation<sup>25</sup>. Moreover, the average cell size (17.5  $\mu\text{m}$  by 6.6  $\mu\text{m}$ ) is comparable to that of *Eosynechococcus grandis*<sup>30</sup>. Other similar structures I have observed have closely spaced transverse markings which are either absent or well defined in individual specimens which taper at both ends (Fig. 4c). Rarer dark coloured and strongly sinuous, filamentous forms, 10–12  $\mu\text{m}$  wide and up to 100  $\mu\text{m}$  long, are constricted towards one of the ends. The ends are otherwise smoothly rounded and conical in shape (Fig. 4b). These structures are

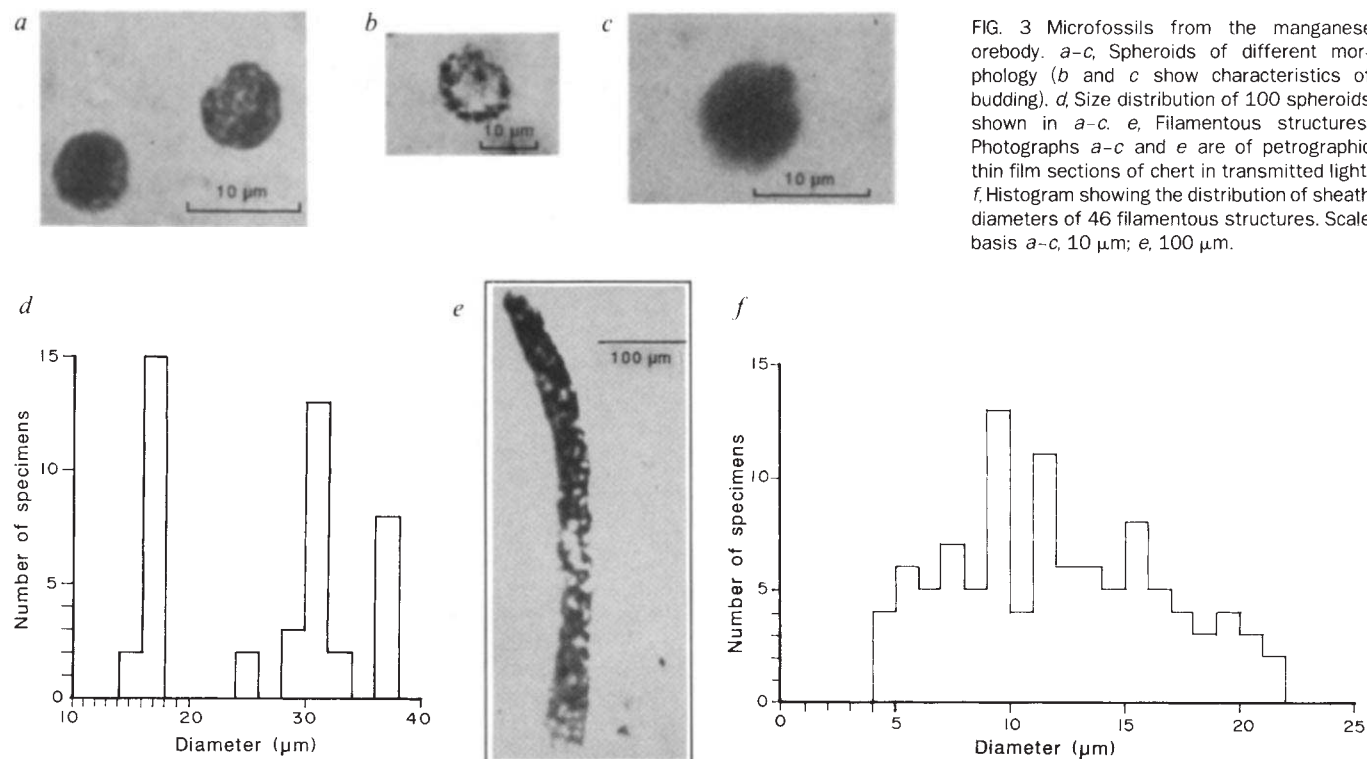
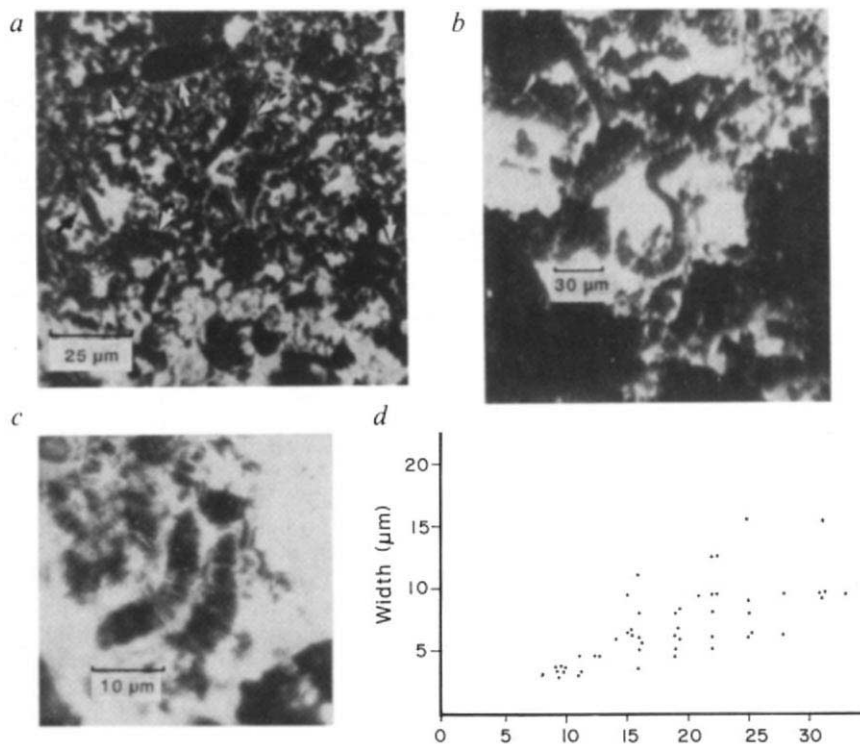


FIG. 3 Microfossils from the manganese orebody. a–c, Spheroids of different morphology (b and c show characteristics of budding). d, Size distribution of 100 spheroids shown in a–c. e, Filamentous structures. Photographs a–c and e are of petrographic thin film sections of chert in transmitted light. f, Histogram showing the distribution of sheath diameters of 46 filamentous structures. Scale basis a–c, 10  $\mu\text{m}$ ; e, 100  $\mu\text{m}$ .

FIG. 4 a–c, Microbial structures from manganese oxide sediments (white) admixed with chert (black). The photographs are of polished samples of manganese ore taken under reflected light in oil medium. d, Size distribution diagram of *Eosynechococcus*-like structures based on 50 individual specimens. Scale basis a, 25  $\mu\text{m}$ , b, 30  $\mu\text{m}$ , c, 10  $\mu\text{m}$ .





clearly of biological origin, and their size and morphology tend to suggest a cyanobacterial affinity.

The occurrence of microfossils within the manganese orebody suggests the diversification of the Proterozoic habitat. It also strengthens the view that Proterozoic microfossils are not confined to early diagenetic chert or shale<sup>31,32</sup>. The average size of the coccoids and filaments are consistent with the Late Proterozoic age determined for the Penganga Group as well as with the observation of a gradual increase in the mean diameter of microfossils at this time<sup>33</sup>. The facies distribution and the sedimentological evidence suggests that the microflora were deposited in deep-water off-shore areas during a marine transgression over an intracratonic shelf. A similar scenario has been suggested for the deposition of microfossils of the Transvaal Supergroup<sup>34</sup> and the Little Belt Supergroup<sup>27</sup>. These, possibly photosynthetic, microorganisms might have acted as oxygen donors, thus facilitating Mn deposition through oxidation of Mn<sup>2+</sup> to Mn<sup>4+</sup> as well as through formation of organo-metallic complexes<sup>1,18</sup>. The intensity of carbonization (yellow to brown colour) together with medium translucency of these microfossils indicate that the diagenetic temperature of the orebody was less than 200 °C<sup>15,35</sup>. Finally, the Penganga microfossils represent the only occurrence of Mn-oxide-chert facies microbiota presently known from India. □

Received 9 March; accepted 3 April 1989.

1. Roy, Supriya, *Manganese Deposits* (Academic, London, 1981).
2. Litherland, M. & Malan, S. P. *J. geol. Soc. Lond.* **129**, 543-544 (1973).
3. Doyen, E. A. in *Ores in Sediments* (eds Amstutz, G. C. & Berand, A. J.) 125-136 (Springer, Berlin, 1973).

4. King, W. *Mem. geol. Soc. India* **18**, 151-311 (1981).
5. Heron, A. M. *J. Hyderabad geol. Surv.* **5**(2), 1-129 (1949).
6. Subba Raju, M. et al. *Rec. Geol. Surv. India* **110**, 39-59 (1978).
7. Roy, Supriya, Bandopadhyay, P. C. & Bose, P. K. in *Conditions of Formation of Ore Deposits. Proc. Sixth IAGOD Symp.* **2**, (eds Janelidze, T., Tvalchrelidze, A. & Sheglov, A.) 787-792 (IAUKA, Moscow, 1986).
8. Chaudhury, A. K. et al. *J. geol. Soc. India* (in the press).
9. Bandopadhyay, P. C. *Miner. Deposita* **23**, 115-122 (1988).
10. Roy, Supriya, Bandopadhyay, P. C. & Bose, P. K. *Twelfth Int. Sédim. Cong. Australia Abstr.* Vol. 8, 263 (1986).
11. Roy, Supriya, Bandopadhyay, P. C., Perceil, E. A. & Fakuoka, M. in *Manganese Metallogenesis Vol. 2 Ore geol. Rev.* (in the press).
12. Barghoorn, W. A. & Tyler, S. A. *Science* **147**, 563-577 (1965).
13. Schopf, J. W. *J. Paleont.* **42**, 651-683 (1968).
14. Licari, G. R. *J. Paleont.* **52**, 762-792 (1978).
15. Oehler, H. *J. econ. Geol.* **72**, 1393-1409 (1977).
16. Awramik, S. M., Golubik, S. & Barghoorn, E. S. *Geol. Soc. Am. Abstr. Prog.* **4**, 438 (1972).
17. Golubic, S. & Barghoorn, E. S. in *Fossil Algae* (ed. Flugel, E.) 1-14 (Springer, Berlin, 1977).
18. Walter, M. R. & Hofmann, H. J. in *Iron Formation: Facts and Problems* (eds Trendall, A. F. & Moriss, R. C.) 373-400 (Elsevier, Amsterdam, 1983).
19. Zhang, Y. *Precambrian Res.* **38**, 165-175 (1988).
20. Schopf, J. W. & Sovietov, Yu. K. *Science* **193**, 143-146 (1976).
21. Oehler, H. *J. Bull. geol. Soc. Am.* **87**, 1143-1152 (1976).
22. Schopf, J. W. *Origins of Life* **7**, 19-36 (1976).
23. Desikachary, T. V. *Cyanophyta* (Indian Council Agric. Res., New Delhi, 1959).
24. Knoll, A. H., Barghoorn, E. S. & Awramik, S. M. *J. Paleont.* **52**, 976-992 (1978).
25. Hofmann, H. J. & Schopf, J. W. in *Earth's Earliest Biosphere: Its Origin and Evolution* (ed. Schopf, J. W.) 321-360 (1983).
26. Knoll, A. H., Strother, P. K. & Susan, R. *Precambrian Res.* **38**, 257-279 (1988).
27. Horodysky, R. J. *J. Paleont.* **54**, 649-663 (1980).
28. Knoll, A. H. & Golubic, S. *Precambrian Res.* **10**, 115-151 (1979).
29. Oehler, H. *J. Geol. Soc. Am. Bull.* **87**, 117-129 (1976).
30. Hofmann, H. J. *J. Paleont.* **50**, 1040-1075 (1976).
31. Schopf, J. W. *A. Rev. Earth planet. Sci.* **3**, 213-250 (1975).
32. Knoll, A. H. *A. Rev. Microbiol.* **39**, 391-417 (1985).
33. Schopf, J. W. *Precambrian Res.* **5**, 143-173 (1977).
34. Klein, C., Beukes, N. J. & Schopf, J. W. *Precambrian Res.* **36**, 81-94 (1987).
35. Sengupta, P. *Rev. Palaeobot. Palynol.* **19**, 173-192 (1975).

ACKNOWLEDGEMENTS. I thank Drs Supriya Roy, Jadupur University, P. C. Banerjee, I.I.C.B., Calcutta for encouragement, A. K. Maitra, Geological Survey of India, for his help, and S. N. Dey, I.I.C.B. for help with photography.

## Facilitation of voltage-gated ion channels in frog neuroglia by nerve impulses

Hector Marrero, Michael L. Astion, Jonathan A. Coles\* & Richard K. Orkand†

Institute of Neurobiology and Department of Physiology, University of Puerto Rico Medical Sciences Campus, 201 Boulevard del Valle, San Juan, Puerto Rico 00901, USA

THE functions of glial cells in the nervous system are not well defined, with the exception of myelin production by oligodendrocytes<sup>1</sup>, uptake of amino-acid synaptic transmitters<sup>2</sup>, and a contribution to extracellular potassium homeostasis<sup>3-5</sup>. Neuroglia have receptors for neurotransmitters<sup>6,7</sup> which may be involved in neuron-glial interactions. Recent studies have demonstrated voltage-gated ion channels in glial membranes<sup>8-12</sup>. In a study of the optic nerve of the frog, small areas of the surface were examined with the loose patch-clamp method<sup>13,14</sup>, and voltage-gated Na<sup>+</sup> and K<sup>+</sup> channels, presumably located in the membranes of the astrocytes forming the glia limitans, were identified<sup>15,16</sup>. We now report that nerve impulses in the axons of the frog optic nerve transiently alter the properties of the voltage-dependent membrane channels of the surface glial cells (astrocytes), a demonstration of a new form of neuron-glial interaction.

Figure 1a is a diagram of the experimental set-up. The distal end of the nerve was held in a suction electrode so that the axons could be stimulated with brief electrical pulses. A loose patch-clamp pipette<sup>14</sup> was applied to the middle of the nerve a few millimetres from the point of stimulation. Figure 1b shows currents recorded during the passage of the nerve impulse as the feedback circuit clamps the electrode potential at zero in

the presence of the field potential. These currents show the time course and relative amplitude of the compound action potential. The same electrode was used to study the electrical properties of the astrocyte membranes. Voltage steps applied to the electrode produced a family of current responses, as described previously<sup>16</sup> (Fig. 1c).

Several arguments support the conclusion that these currents were flowing across the membranes of the astrocytes, whose endfeet form a layer up to a few micrometres thick on the surface of the nerve, and thus are the membranes closest to the micropipette. The currents have a complex, but reproducible, dependence on voltage and time. The initial inward component of the current is abolished by tetrodotoxin (0.4 µM), and the fast outward current by 4-aminopyridine (0.5 mM)<sup>16</sup>. Thus, the currents result from voltage steps across a cell membrane rather than some external resistance. Further results support the conclusion that these membrane currents originated in the surface glia. First, when Na<sup>+</sup> was removed from both the bath and the patch electrode no voltage-dependent inward currents were recorded. If only the pipette was filled with normal Ringer's solution, so that Na<sup>+</sup> was high only at the glia limitans, inward currents were recorded. Second, in an experiment suggested by W. Stühmer, when [K<sup>+</sup>] in the bath and the electrode were increased to 30 mM (10× normal), the axons became inexcitable. If the holding potential was made positive (so as to polarize the underlying membrane), however, inward currents were again obtained. Third, when Ba<sup>2+</sup> (5 mM) was added to the bath the nerve impulse was little affected, but the current-voltage curve obtained from the patch recording was shifted as expected if the glial cells were depolarized, which is known to be the case<sup>17</sup>. Fourth, currents of comparable time course and about 10% of control amplitude were recorded from the surface of the optic nerve 4 months after retinal ablation when all the axons had degenerated.

Our principal result in itself provides evidence that the patch-clamp currents are from the glia and not from the axons. For depolarizing steps between 40 and 100 mV, the amplitudes of the inward and outward currents were increased following nerve

\* Permanent address: INSERM, U176, rue Camille Saint-Saëns, F33077 Bordeaux, France.

† To whom correspondence should be addressed.

stimulation. Figure 2a illustrates this finding for a voltage step of 50 mV. Traces 1 and 3 are control current responses. Trace 2 shows how the amplitudes of both the inward and outward currents were increased when a nerve volley had passed 15 ms earlier. We recorded current-voltage curves with and without a nerve volley. The nerve volley, delivered 50 ms before the voltage step, produced a shift in the current-voltage curve of 23 mV (s.d. = 4.7,  $N = 5$ ) for a voltage step of 60 mV. This shift seems to account for the change in shape of the currents. The nerve

volley caused no change in the patch seal resistance, and the results were not affected by pressing the electrode more or less firmly against the nerve. This facilitation of the patch currents is the opposite of the changes at this time in the axons, which were relatively refractory (Fig. 2b). When the nerve was stimulated by two supramaximal stimuli separated by 15 ms, the response to the second stimulus was smaller. This was because a fraction of the  $\text{Na}^+$  channels was inactivated, so some axons did not conduct, and those that did, conducted more slowly<sup>18,19</sup>.

Why are the glial currents facilitated by neuronal activity? The most obvious possibility is that the facilitation is due to glial depolarization resulting from an accumulation of  $\text{K}^+$  which is released from the axons into the narrow extracellular clefts during the nerve impulse<sup>5,20</sup>. There are two reasons to believe that the facilitation is not directly related to the  $\text{K}^+$ -induced depolarization. First, facilitation is still observed, although in a slightly altered fashion, in the presence of external  $\text{Ba}^{2+}$  (2 mM), which is known to block the glial depolarization induced by a rise in extracellular  $[\text{K}^+]$  (ref. 17). Second, the time course of the decay of facilitation and the time course of the decay of the glial depolarization differ markedly. Figure 3a shows that the facilitation of the glial currents following either a single nerve volley or a train of 10 stimuli reached a peak about 30 ms after the end of stimulation and then decayed. Figure 3b shows that, for a train of 10 nerve impulses, the glial depolarization was still present (half-time for decay, 1.43 s) after the facilitation of the glial currents had disappeared (half-time, 0.23 s). Thus, facilitation is not simply related to membrane depolarization.

Another factor, besides the  $\text{K}^+$ -induced depolarization, which might be responsible for the facilitation of the glial currents is a fall in extracellular  $\text{Ca}^{2+}$  concentration. A rapid transient

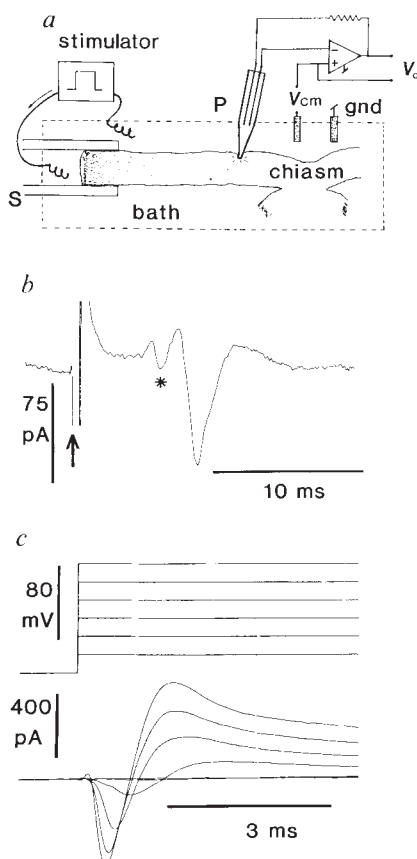


FIG. 1 *a*, Experimental set-up. The distal end of the optic nerve was held in a suction electrode (S) to stimulate axonal action potentials. A Ringer's-filled glass pipette electrode (P) was placed on the glial surface and connected to a loose patch-clamp amplifier. *b*, Compound action potential. The axons were stimulated supramaximally (arrow). First deflection (\*) is the action potential in the small number of myelinated axons, and the second, larger current is from slowly conducting unmyelinated axons. *c*, Glial currents induced by depolarizing steps. Top: pipette potential decreased from -20 to -120 mV. Currents shown are the sum of symmetrical depolarizing and hyperpolarizing steps (average of three repetitions). No voltage-dependent current was observed with steps of -20 and -40 mV. With increasing depolarization an inward current (downwards) appears, followed by an outward current.

**METHODS.** The optic nerve (1-mm diameter), a central nervous system pathway, was dissected from frogs (*Rana pipiens*, 5 cm body length). It contains axons, ~3% of which are myelinated by oligodendrocytes, and astrocytes which form the glia limitans. Ringer's solution had the following composition (mM): NaCl (110), KCl (2),  $\text{CaCl}_2$  (1.9), HEPES (5), pH 7.4. Recording electrodes (10–25  $\mu\text{m}$ , 0.1–0.3 Mohm) were filled with bath solution and connected to the loose patch-clamp amplifier (Stühmer Elektronik)<sup>15</sup> through silver wires freshly coated with molten AgCl. They were gently pressed on the nerve until a seal formed with a resistance equal to that of the pipette. Symmetrical voltage steps were applied to the electrode, and the resulting current flow was measured and corrected for leakage and capacitive artefacts. The partially corrected signal was digitized (IBM XT computer) and further analysed using pCLAMP (version 3.02, Axon Instruments). Glial membrane potential was measured with conventional 3M KCl microelectrodes.

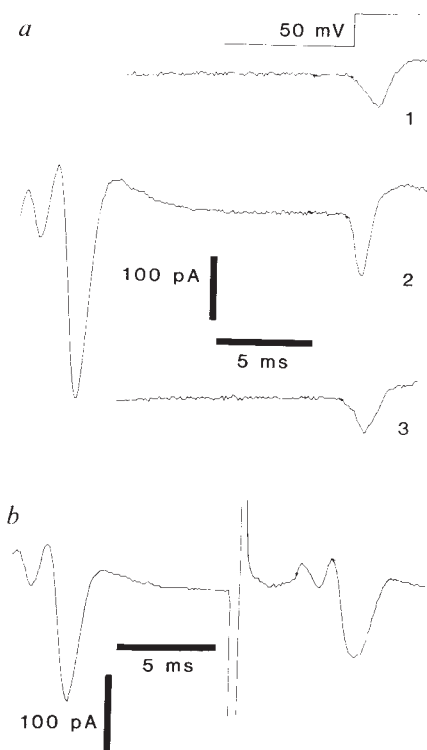


FIG. 2 *a*, Effect of nerve impulses on voltage-dependent currents in glial cells. 1, Glial current in response to a depolarizing step of 50 mV (average of three sweeps as in Fig. 1c). 2, A single nerve volley precedes the same voltage step by 15 ms and the glial current is facilitated. 3, Control taken 1 min after the nerve stimulus was turned off. *b*, Recording of compound action potentials to successive nerve volleys. When the nerve is stimulated twice at an interval of 15 ms the response to the second stimulus is reduced, that is, the nerve is in the relative refractory period.

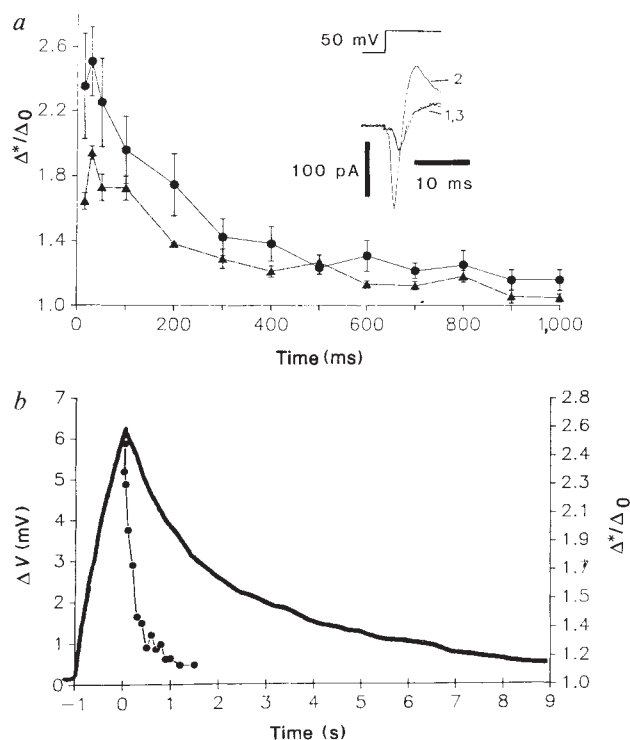


FIG. 3 Time course of facilitation of glial currents produced by nerve impulses. *a*, Ratio of glial current amplitudes (peak-to-peak)  $\Delta^*/\Delta_0$  as a function of time after a single nerve volley (▲), or after a train of nerve impulses at 10 Hz for 1 s (●). Inset shows the superimposed glial currents: 1, control; 2, 50 ms after the train; 3, one minute after the train. The currents are facilitated by the impulses and return to control amplitude in about 1 s. Bars show s.d.;  $N = 4$  (▲),  $N = 12$  (●). *b*, Time course of facilitation compared with the depolarization of the glial membrane produced by the nerve impulses. The filled circles indicate the facilitation following a train of impulses at 10 Hz for 1 s (same data as in Fig. 3*a*). The heavy line is a record of the time course of the glial depolarization ( $\Delta V$ ) following an identical train of impulses, recorded from a glial cell with an intracellular electrode in a separate experiment.

decrease in  $[Ca^{2+}]$  following impulse activity, which has been observed in other tissues<sup>21</sup>, would facilitate the inward  $Na^+$  current because of an alteration in membrane surface charge<sup>22</sup>. Finally, the facilitation might be due to the release of an unidentified substance from the axons.

Our finding that the behaviour of glial membrane channels is altered by nerve activity demonstrates a possible form of signalling between neurons and glia. The mechanism and functional role of this signalling remain to be clarified. □

17. Aston, M. L., Obaid, A. L. & Orkand, R. K. *J. gen. Physiol.* (in the press).
18. Erlanger, J. & Gasser, H. S. *Electrical Signs of Nervous Activity* (Univ. Pennsylvania Press, Philadelphia, 1937).
19. Bishop, G. H. *Am. J. Physiol.* **106**, 460–474 (1933).
20. Ransom, B. R., Yamate, C. L. & Connors, B. W. in *Ion Measurements in Physiology and Medicine* (eds Kessler, M., Höper, J. & Harrison, D. K.) 206–213 (Springer, Berlin, 1985).
21. Pumain, R., Kurcewicz, I. & Louvel, J. *Science* **222**, 177–179 (1983).
22. Frankenhaeuser, B. & Hodgkin, A. L. *J. Physiol., Lond.* **137**, 218–244 (1957).

ACKNOWLEDGEMENTS. We thank W. Stühmer, J. del Castillo, I. Hentall and C. Zuazaga for suggestions and P. M. Orkand for unpublished electron micrographs of the surface of the optic nerve. This work was supported by the NIH and NSF.

## Multiple conductance channels in type-2 cerebellar astrocytes activated by excitatory amino acids

Maria M. Usowicz\*, Vittorio Gallo\* & Stuart G. Cull-Candy†

MRC Receptor Mechanisms Research Group, Department of Pharmacology, University College London, Gower Street, London WC1E 6BT, UK

L-GLUTAMATE and L-aspartate are thought to have a widespread function as synaptic transmitters in the mammalian central nervous system and there are at least three types of neuronal glutamate receptors, which can be activated by the selective agonists *N*-methyl-D-aspartate (NMDA), quisqualate and kainate<sup>1,2</sup>. Recent experiments indicate that glutamate receptors also occur in astrocytes<sup>3–6</sup>. We have used patch-clamp methods<sup>7</sup> to determine whether one type of macroglial cell, the type-2 astrocyte, possesses glutamate receptors, as previously proposed from neurochemical studies<sup>5</sup>. We find that glutamate and related amino acids can evoke whole-cell and single-channel currents in type-2 astrocytes from rat cerebellum. Although these cells are found mainly in white matter<sup>8–10</sup>, where neurotransmission does not occur, their processes are closely associated with axons at nodes of Ranvier<sup>9,10</sup>, suggesting that such receptors are involved in neuronal–glial signalling at the node. Our experiments show that glial cells possess quisqualate- and kainate-receptor channels but lack receptors for NMDA. Interestingly, these glutamate channels exhibit multiple conductance levels that are similar in amplitude to the neuronal glutamate channels<sup>11,12</sup>.

Application of L-glutamate (30  $\mu$ M), quisqualate (30  $\mu$ M), or kainate (30–100  $\mu$ M), to cells in whole-cell voltage clamp, evoked a current change accompanied by a clear increase in noise level (Fig. 1). The relationship between current amplitude and membrane potential (Fig. 1*b*) gave a reversal potential of  $+5.9 \pm 1.1$  mV (12 cells, pooling data for glutamate, quisqualate and kainate). The fact that these currents reversed near 0 mV, and showed an increase in noise, implies that they resulted from the opening of ion channels (as described for neuronal glutamate-receptor channels<sup>13–17</sup>). Furthermore, the kainate responses were reduced by  $\sim 70\%$  with the glutamate-receptor antagonist kynurenic acid (100  $\mu$ M).

It has been reported that glutamate, quisqualate and kainate produce similar whole-cell responses in rat cerebral astrocytes<sup>6</sup>, although the type of astrocyte was not defined as type-1 or type-2. By contrast, the whole-cell glutamate current in glial cells possessing an electrogenic glutamate-uptake carrier (retinal Muller cells<sup>18</sup> and cerebellar type-1 astrocytes<sup>19</sup>), is not associated with a change in noise and remains inward-going at potentials as positive as +80 mV. It seems therefore that the glutamate current in type-2 astrocytes arises mainly from direct activation

Received 31 January; accepted 30 March 1989.

1. Bunge, R. P. *Physiol. Rev.* **48**, 197–251 (1968).
2. Ransom, B. R. & Carlini, W. G. in *Astrocytes*, (eds Vernadakis, A. & Fedoroff, S.) Vol. 2, 1–49 (Academic, Orlando, 1986).
3. Orkand, R. K. *Ann. N.Y. Acad. Sci.* **481**, 269–272 (1986).
4. Syková, E. *Prog. Biophys. molec. Biol.* **42**, 135–189 (1983).
5. Orkand, R. K., Nicholls, J. G. & Kuffler, S. W. *J. Neurophysiol.* **29**, 788–806 (1966).
6. de Vellis, J., Inglish, D. & Galey, F. in *Cellular Aspects of Neural Growth and Differentiation* (ed. Pease, D.) 23–32 (UCLA Forum in Medical Sciences, **14**, Univ. California Press, 1971).
7. *Glial Cell Receptors* (ed. Kimelberg, H. K.) (Raven, New York, 1988).
8. MacVicar, B. A. *Science* **226**, 1345–1347 (1984).
9. Newman, E. A. *Nature* **317**, 809–811 (1985).
10. Chiu, S. Y., Shrager, P. & Ritchie, J. M. *Nature* **311**, 156–157 (1984).
11. Barres, B. A., Chun, L. L. Y. & Corey, D. P. *Glia* **1**, 10–30 (1988).
12. Bevan, S., Chiu, S. Y., Gray, P. T. A. & Ritchie, J. M. in *Ion Channels in Neural Membranes* (eds Ritchie, J. M., Keynes, R. D. & Bolis, L.) 159–174 (Liss, New York, 1986).
13. Almers, W., Stanfield, P. R. & Stühmer, W. *J. Physiol., Lond.* **336**, 261–284 (1983).
14. Stühmer, W., Roberts, W. M. & Almers, W. in *Single Channel Recording* (eds Sakmann, B. & Neher, E.) 123–132 (Plenum, New York, 1983).
15. Konnerth, A., Orkand, P. M. & Orkand, R. K. *Glia* **1**, 225–232 (1988).
16. Ortiz, S., Rodriguez, O., Orkand, P. M., Orkand, R. K. & Marrero, H. *Puerto Rico Health Sciences Journal* **7**, 141–143 (1988).

\* Present addresses: Department of Pharmacology, University of Bern, Friedbühlstrasse 49, Bern 3010, Switzerland (M.M.U.); Laboratory of Physiopathology, Istituto Superiore di Sanità, Viale Regina Elena, 299, 00161 Rome, Italy (V.G.).

† To whom correspondence should be addressed.



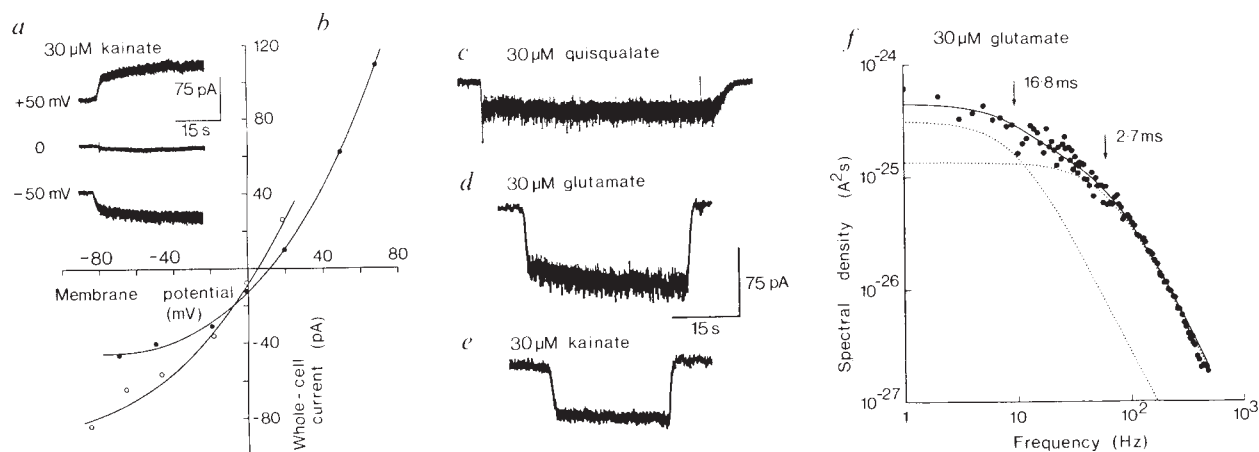


FIG. 1 Whole-cell currents evoked by glutamate-receptor agonists in cerebellar type-2 astrocytes. *a*, Currents produced by steady perfusion of 30  $\mu$ M kainate onto a cell held at -50 mV, 0 mV and +50 mV. *b*, Relationship between current and membrane potential.  $\bullet$ , 30  $\mu$ M kainate (same cell as in *a*);  $\circ$ , 30  $\mu$ M quisqualate (same cell as in *c*). Reversal potentials are 11.5 mV, kainate; 5.5 mV, quisqualate. *c*, Currents evoked by 30  $\mu$ M quisqualate (-70 mV). *d*, 30  $\mu$ M glutamate (-70 mV). *e*, 30  $\mu$ M kainate (-70 mV, same cell as *d*). Calibration *a*, *c*, *d*, *e*: 75 pA and 15 s. *f*, Spectral density of glutamate-noise (-90 mV), fitted with the sum of two Lorentzian components (continuous line). Individual components indicated by dashed lines. Arrows indicate the time constant,  $\tau$ , of each component (according to  $\tau = 1/(2\pi f_c)$ , where  $f_c$  is the cut-off frequency of each component):  $\tau_1 = 16.8$  ms and  $\tau_2 = 2.7$  ms. The estimated single-channel conductance is 7.9 pS, which is in the range previously obtained from quisqualate and kainate noise in neurons. Points above 60 Hz have been averaged.

**METHODS.** Cerebellar type-2 astrocyte-enriched cultures were obtained from 8-day-old rats, as previously described (see ref. 5). Cerebellar cells were dissociated with trypsin and DNase and plated at low density on poly-L-lysine ( $10 \mu\text{g ml}^{-1}$ ;  $M_r = 53,000$ ) coated coverslips. Cells were cultured in Eagle's basal medium with 10% fetal calf serum. Because type-2 astrocytes divide infrequently in culture<sup>4</sup> they are quickly overgrown by rapidly dividing type-1 astrocytes in higher density cultures (of the sort previously

used in ref. 19) and in older cultures. Recordings were therefore made after 3–5 days *in vitro* from type-2 astrocytes that were distinguished morphologically from the few oligodendrocytes, type-1 astrocytes and granule cells present. Type-2 astrocytes were stellate shaped with dark cell bodies and tended to occur in small groups; these could be readily identified under phase contrast optics ( $\times 640$  water immersion). Oligodendrocytes had bright cell bodies with a fine network of radial processes, whereas type-1 astrocytes possessed a polygonal, epithelioid morphology with a flattened appearance. The morphological identification was verified by indirect immunofluorescence (see refs 8 and 5 for references) using other cover slips from the same culture. Labelling with anti-glial fibrillary acid protein (GFAP) antibodies, anti-galactocerebroside antibodies (donated by Barbara Ranscht) and the monoclonal antibodies LB1 (donated by Jim Cohen & Graham Wilkin) and A2B5, indicated the presence of both GFAP-positive/A2B5-positive/LB1-positive cells (type-2 astrocytes), and galactocerebroside positive cells (oligodendrocytes), in a 9:1 ratio. The bathing solution for patch-clamp experiments was, in mM: NaCl, 145; KCl, 2.8;  $\text{CaCl}_2$ , 1; Na-HEPES, 10; glucose, 5; pH 7.2 with NaOH. The pipette (intracellular) solution contained, in mM: CsF, 110; CsCl, 30 (or CsCl, 140), NaCl, 4;  $\text{K}_2\text{-EGTA}$ , 5; K-HEPES, 10;  $\text{CaCl}_2$ , 0.5; pH 7.2 with KOH. Drugs were applied in the bathing medium by perfusion through a fine plastic tube placed beneath the water immersion objective. Temperature = 20–24  $^{\circ}\text{C}$ .

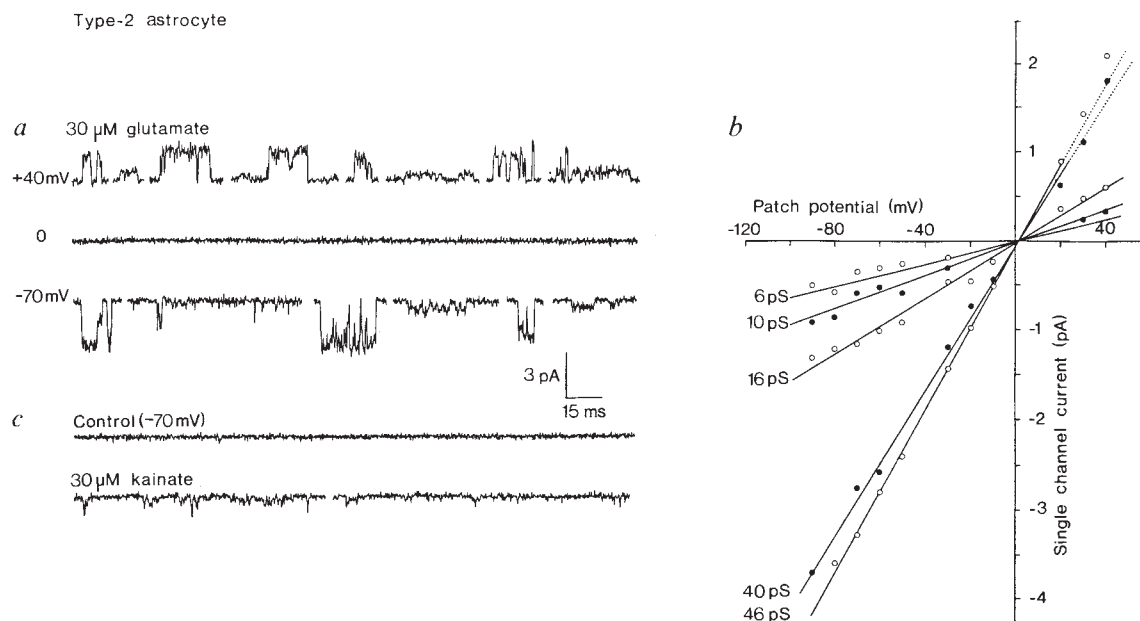


FIG. 2 Single-channel currents produced by glutamate and kainate in an outside-out patch of a cerebellar type-2 astrocyte. *a*, Multiple-amplitude currents produced by 30  $\mu$ M glutamate are inward at -70 mV and outward at +40 mV; the reversal potential is near 0 mV. *b*, Relationship between multiple current amplitudes and patch potential for channels activated by glutamate in *a*. The mean reversal potential is +1.5 mV, and each line has been drawn through this value with the slope calculated from least squares fitting. The points, at +40 mV, belonging to the two largest levels (dotted lines), deviate from the fitted lines which have slope conductances of 46, 40, 16, 10 and 6 pS. For outward currents at +40 mV, the two largest conductance levels are 54 pS and 47 pS. The linear relationship in the potential range -90 to +20 mV, however, indicates that the slight outward rectification of whole-cell currents (Fig. 1*b*) is not due to non-ohmic behaviour of single-channel conductance. It is likely that the channels possess more than five conductance levels because the smallest level (6 pS), when

activated by glutamate or quisqualate (Fig. 3), seemed to represent the mean of 3 and 8 pS levels. These were readily identified when connected by a direct transition, at extremes of potential (more negative than -70 mV). 'Noisy' openings, intermediate to the conductances of 40 and 16 pS, or direct transitions to these levels from other open levels, were also occasionally observed (not shown). *c*, Upper trace, control noise level of the patch after wash out of glutamate (-70 mV). Lower trace, single-channel currents produced by 30  $\mu$ M kainate in the same patch cannot be individually resolved, but appear as a noise increase. All single channel currents in this study were recorded in patches isolated from the cell body, and all patches tested with glutamate gave responses. Patch-pipettes were made from hard pyrex thick walled (Clark's Electromedical) glass, and were coated with Sylgard resin. Calibration for *a* and *c*: 3 pA and 15 ms. Currents filtered at 1.56 kHz, -3 dB.

of ion channels, whereas in type-1 astrocytes it results predominantly from the activation of an uptake mechanism. The presence of a low density of receptor-channels in type-1 astrocytes, however, cannot be excluded.

Type-2 astrocytes did not respond to 30–100  $\mu$ M NMDA (10 cells) or 30–100  $\mu$ M aspartate (8 cells), regardless of the presence of glycine<sup>20</sup>, although these same cells were sensitive to the other agonists. Because our experiments were made in nominally  $Mg^{2+}$ -free solutions, the lack of NMDA-sensitivity indicates an absence of NMDA receptors rather than  $Mg^{2+}$ -block of the channel, of the sort described in neurons<sup>13,21</sup>.

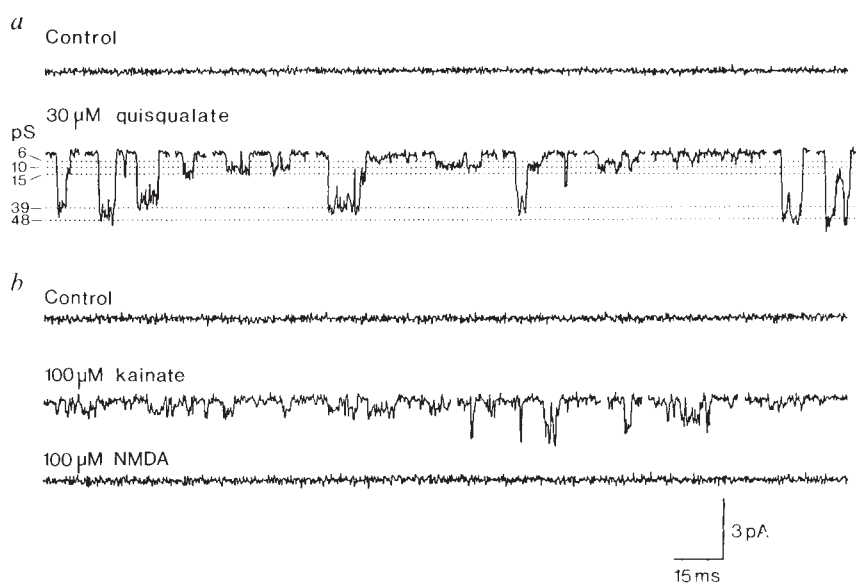
From a spectral analysis of whole-cell noise, an estimate of the channel characteristics was made (see Fig. 1f). In addition, the single-channel currents were examined directly in outside-out patches. As illustrated in Fig. 2a, 30  $\mu$ M glutamate (examined in five patches) invariably produced multiple conductance openings. All these conductances showed a linear dependence on patch potential between –90 and +20 mV (see

Fig. 2b). The multiple conductance levels activated by quisqualate (19 patches; Fig. 3a), resembled those produced by glutamate: pooled single-channel data for glutamate and quisqualate, measured at a minimum of five potentials, gave mean-slope conductances of  $47 \pm 0.5$ ,  $39 \pm 1.1$ ,  $15 \pm 0.3$ ,  $10 \pm 0.3$  and  $6 \pm 0.3$  pS ( $\pm$ s.e.m.; 6 patches), with a mean reversal potential of  $+3.5 \pm 0.8$  mV. Most of the quisqualate openings ( $\sim 80\%$ ) were below 30 pS, and direct transitions between levels were relatively infrequent (Fig. 3).

In some patches, kainate (30–100  $\mu$ M) produced small brief events which appeared as a noise increase (Fig. 2c) whereas in others it gave multiple conductance openings, with values similar to quisqualate openings of less than 30 pS, superimposed on the noise increase (Fig. 3b). Direct transitions between multiple kainate levels were also rare. Unlike glutamate and quisqualate, however, kainate was highly selective for levels below 30 pS. Spectra of kainate currents in outside-out patches were fitted with two lorentzian curves (not shown), with slow and fast time

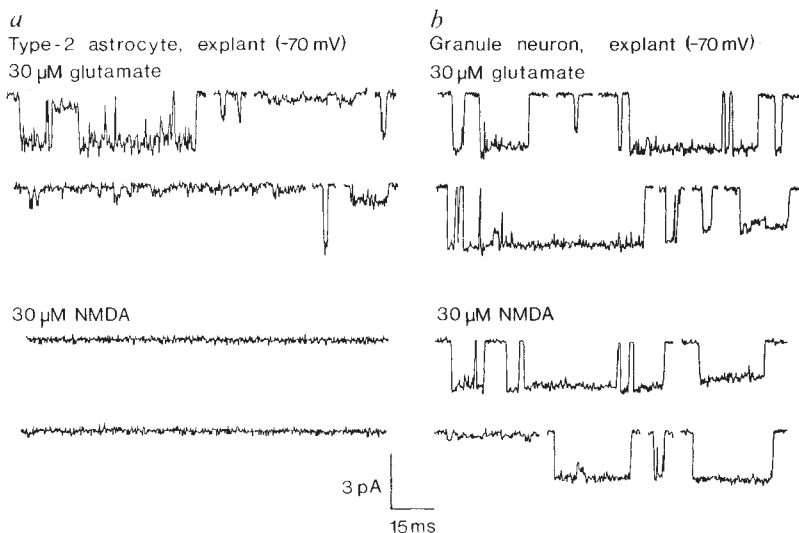
FIG. 3 Single-channel currents produced in outside-out patches (–70 mV) of cerebellar type-2 astrocytes by the selective glutamate receptor agonists quisqualate, kainate and NMDA. a, Upper trace: control noise level. Lower trace:

current produced by 30  $\mu$ M quisqualate show multiple current amplitudes (indicated by dotted lines). These correspond to slope conductances of 48, 39, 15, 10 and 6 pS (interpolated reversal potential +1.9 mV). Values for the largest level were increased at potentials more positive than +20 mV (not shown), as for glutamate channels in Fig. 2b. It is not clear whether the quisqualate currents reflect activation of several channels or one channel, but direct transitions between levels were infrequent. However, transitions were observed between levels above and below 30 pS, which is inconsistent with the possibility that levels above and below 30 pS arose from separate channels. b, Upper trace: control noise level. Middle trace: kainate (100  $\mu$ M) produced openings with directly resolvable multiple levels, and concomitantly activated conductances that are too small to be resolved individually, but appear as a noise increase. Lower trace: the same patch gave no currents in response to 100  $\mu$ M NMDA. Slope conductance estimates for the discernible levels were 19, 12 and 8 pS, with a reversal potential of –2.4 mV. Openings to these levels were often briefer in duration than the currents illustrated, and on average briefer than openings produced in the same patches by quisqualate (compared in two patches) or glutamate (compared in one patch). The high selectivity of kainate for levels below 30 pS, and the high frequency of activation of small non-resolvable conductances is reflected in the fact that, for similar mean currents, whole-cell kainate noise was smaller than quisqualate noise (see Fig. 1). The size of whole-cell glutamate



noise was intermediate between values for kainate and quisqualate, indicating the ability of glutamate to activate a mixture of kainate and quisqualate channels. Calibration: 3 pA and 15 ms. Currents filtered at 1.56 kHz (3 dB).

FIG. 4 a, The effect of glutamate and NMDA on an outside-out patch from a cerebellar type-2 astrocyte in explant culture (see ref. 19); cells not exposed to enzymes during preparation of the culture, unlike the cells represented in Figs 1–3. Glutamate (30  $\mu$ M, upper traces) produced multiple-amplitude openings (mainly less than 30 pS), whereas NMDA (30  $\mu$ M, lower traces) had no effect (1  $\mu$ M glycine and 1  $\mu$ M strychnine present in bathing solution<sup>20</sup>). A lack of NMDA-sensitivity was found in eight patches from type-2 astrocytes in explant culture, whereas the currents produced in these same patches by 30  $\mu$ M glutamate (seven patches) and 100  $\mu$ M kainate (two patches) were qualitatively similar to those recorded in patches from enzyme treated cells. b, By contrast, a patch from a cerebellar granule neuron in explant culture showed single-channel currents in response to both glutamate (upper traces) and NMDA (lower traces). Calibration: 3 pA and 15 ms. Most openings are  $>30$  pS, consistent with the view that, although glutamate activates non-NMDA channels in type-2 astrocytes, it preferentially activates NMDA-channels in granule cells<sup>19</sup>, as in other types of neurons<sup>11,12,13,19</sup>. The currents in b are not typical of all types of neurons, and the noise level of glutamate-evoked single-channel currents differs among large cerebellar neurons<sup>17</sup> cerebellar granule neurons, and cerebellar type-2 astrocytes. Furthermore, the duration of openings  $>30$  pS produced by quisqualate in cerebellar type-2 astrocytes is on average shorter than NMDA-evoked openings in large cerebellar neurons and cerebellar granule neurons. The quisqualate 45 pS openings are not blocked by  $Mg^{2+}$ , unlike NMDA 45 pS openings in neurons. We also find that openings  $>30$  pS produced in large cerebellar neurons by 10  $\mu$ M quisqualate are briefer than those produced by 50  $\mu$ M quisqualate<sup>11,17</sup>. This comparison suggests the possibility that, at



least in large cerebellar neurons, high concentrations of quisqualate activate the NMDA receptor, as well as the quisqualate receptor to give a mixture of  $>30$  pS openings differing in duration, sensitivity to  $Mg^{2+}$  block, and noise level.

constants of 7–14 and 0.4–1 ms, ( $V_m = -90$  to  $-70$  mV), respectively, and the estimated values for the single channel conductance were 2–3 pS.

As shown in Fig. 3, the absence of NMDA channels was confirmed in isolated patches exposed to NMDA (30  $\mu$ M, four patches) or aspartate (30–100  $\mu$ M, two patches); the same patches invariably gave currents with glutamate, quisqualate or kainate (Fig. 3b). Furthermore, the possibility that trypsin treatment of cells<sup>5</sup> caused a selective loss of NMDA receptors<sup>22</sup> was excluded by experiments such as that depicted in Fig. 4a, showing that type-2 astrocytes in mechanically dissociated explant cultures<sup>19</sup> were also insensitive to NMDA. Figure 4 illustrates a direct comparison of the channels activated in granule neurons and in type-2 astrocytes, present in the same explant cultures. In the granule cells, NMDA and glutamate produced currents that opened predominantly to levels above 30 pS. Although  $Mg^{2+}$  blocks the NMDA-activated 45-pS conductance in neurons<sup>13</sup>, in cerebellar type-2 astrocytes  $Mg^{2+}$  (0.5 mM) did not induce a flickery block of any of the multiple levels (including the 45-pS levels) activated by quisqualate (eight patches).

Our results demonstrate that type-2 astrocytes have both quisqualate- and kainate-activated channels. In certain respects these are clearly similar to neuronal glutamate channels<sup>11,12</sup> but the data also suggest some differences. For example, certain neurons lacking NMDA receptors fail to give 45-pS openings in response to quisqualate<sup>23</sup>. By contrast, in the astrocytes, 45-pS events were readily obtained with quisqualate, despite the absence of NMDA receptors. Moreover, the fact that  $Mg^{2+}$  blocks NMDA 45-pS events in neurons<sup>13</sup> but not the quisqualate 45-pS levels in astrocytes, raises the intriguing possibility that some quisqualate 45-pS openings may be  $Mg$ -insensitive in neurons (see also Fig. 4 legend). Kainate currents in type-2 astrocytes differed from those described in some neurons<sup>11,12</sup>, in the lack of openings above 30 pS. This may reflect differences in the kainate channels present in the two cell types, or it may indicate that in neurons kainate events above 30 pS are produced by the activation of NMDA<sup>11</sup> or quisqualate receptors. As well as possessing amino-acid receptors, type-2 astrocytes show some other 'neuronal' characteristics both in their antigenic properties<sup>24,25</sup> and in their voltage-activated channels<sup>26</sup>. Our results further suggest that a proportion of glutamate receptors, or messenger (m) RNA encoding these receptors, derived from whole brain tissue may originate from glia. This could be one reason why mRNA injected into *Xenopus* oocytes induces a low density of NMDA receptors<sup>27,28</sup>, compared with quisqualate and kainate receptors.

What is the functional significance of the astrocytic glutamate receptors? The type-2 astrocytes possess end-feet, in close apposition to the nodes of Ranvier *in vivo*<sup>9,10</sup>, that are well positioned to be activated by any axonal release of transmitter. Such release of glutamate from axons has previously been reported<sup>29,30</sup>. Our finding that the receptors remain active in the continued presence of glutamate suggests that any axonally released transmitter acting on nodal astrocytes would cause a steady change in the local ionic environment, and in electrical excitability of the axon. In addition, the ability of type-2 astrocytes to 'detect' or respond to glutamate may be important in the formation or maintenance of the node; such responses may also play a role in the interactions between axons and migrating progenitors of the type-2 astrocytes, which also seem to possess glutamate receptors<sup>5</sup>. □

8. Wilkin, G. P. & Levi, G. in *Astrocytes*, Vol. 1, 245–268 (Academic, London, 1986).
9. French-Constant, C. & Raff, M. C. *Nature* **323**, 335–338 (1986).
10. Miller, R., Fulton, Barbara P. & Raff, M. C. *Eur. J. Neurosci.* **1**, 171–180 (1989).
11. Cull-Candy, S. G. & Usowicz, M. M. *Nature* **325**, 525–528 (1987).
12. Jahr, C. E. & Stevens, C. F. *Nature* **325**, 522–525 (1987).
13. Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. *Nature* **307**, 462–465 (1984).
14. Ishida, A. T. & Neyton, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1837–1841 (1985).
15. Cull-Candy, S. G. & Usowicz, M. M. *Brain Res.* **402**, 182–187 (1987).
16. Ascher, P. & Nowak, L. *J. Physiol. Lond.* **399**, 207–226 (1988).
17. Cull-Candy, S. G. & Usowicz, M. M. *J. Physiol. Lond.* (in the press).
18. Brew, H. & Attwell, D. *Nature* **327**, 707–709 (1987).
19. Cull-Candy, S. G., Howe, J. R. & Ogden, D. C. *J. Physiol. Lond.* **400**, 189–222 (1988).
20. Johnson, J. W. & Ascher, P. *Nature* **325**, 529–531 (1987).
21. Mayer, M. L., Westbrook, G. L. & Vyklicky, L., Jr. *J. Neurophysiol.* **60**, 645–663 (1988).
22. Akaike, N., Kaneda, M., Hori, N. & Krishtal, O. A. *Neurosci. Lett.* **87**, 75–79 (1988).
23. Llano, I., Marty, A., Johnson, J. W., Ascher, P. & Gähwiler, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3221–3225 (1988).
24. Raff, M. C., Miller, R. H. & Noble, M. *Nature* **303**, 390–397 (1983).
25. Levi, G., Gallo, V. & Ciotti, M. T. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1504–1508 (1986).
26. Barres, B. A., Chun, L. L. Y. & Corey, D. P. *Glia* **1**, 10–30 (1988).
27. Gundersen, C. B., Miledi, R. & Parker, I. *Proc. R. Soc. B221*, 127–143 (1984).
28. Verdoorn, T. A., Kleckner, N. W. & Dingledine, R. *Science* **238**, 1114–1116 (1988).
29. Weinreich, D. & Hamerslag, R. *Brain Res.* **84**, 137–142 (1975).
30. Abbott, N. J., Hassan, S. & Lieberman, E. M. *J. Physiol. Lond.* **398**, 63P (1988).

ACKNOWLEDGEMENTS. We thank Corinne Symonds for help with cell culture and antibody labelling, Alistair Mathie, David Colquhoun, Barbara Fulton and Martin Raff for helpful comments and discussions. This work was supported by the Wellcome Trust, the MRC and EMBO (V.G.).

## Interleukin-2 production used to detect antigenic peptide recognition by T-helper lymphocytes from asymptomatic HIV-seropositive individuals

Mario Clerici\*, Naomi I. Stocks\*, Robert A. Zajac†, R. Neal Boswell†, Denise C. Bernstein\*, Dean L. Mann‡, Gene M. Shearer\* & Jay A. Berzofsky§

\* Experimental Immunology Branch, † Viral Carcinogenesis Branch and § Metabolism Branch, National Cancer Institute, National Institutes of Health, Bethesda Maryland 20892, USA

‡ HIV Unit/SGHMMM, Wilford Hall, Lackland AFB, Texas 28236, USA

T LYMPHOCYTES from mice<sup>1</sup> and healthy humans<sup>2</sup> immunized against the human immunodeficiency virus (HIV) envelope have recently been shown to recognize two antigenic regions of the gp160 HIV-envelope protein which have been located on the basis of amphipathicity<sup>3–6</sup>. In HIV-infected humans, T-cell proliferative responses are lost soon after infection<sup>7,8</sup>. Here we demonstrate that interleukin-2 production is often retained even when proliferative activity is absent, and that it can be used to monitor T-helper cell responses by HIV-seropositive donors. We use this approach to investigate the T-helper cell response of 42 asymptomatic HIV-seropositive patients to four synthetic gp160 peptides and to influenza A virus, an antigen requiring intact CD4 T-helper cell function. As many as 67% of the HIV-seropositive donors who retain responsiveness to influenza A virus respond to a single peptide, and 85–90% responded to at least one of the peptides.

We tested the ability of peripheral blood leukocytes (PBL) from asymptomatic, HIV seropositive (HIV<sup>+</sup>), Walter Reed Stage 1 and 2 patients (all of whom had  $>400$  CD4<sup>+</sup> cells mm<sup>-3</sup>) (ref. 9) to produce interleukin-2 (IL-2) when stimulated *in vitro* with influenza-A virus (FLU), or with four HIV envelope peptides: T1, T2, TH4.1, and P18 (Fig. 1; Table 1). The PBL from an HIV<sup>-</sup> donor (Fig. 1a) generated a strong IL-2 response to FLU, but failed to respond to any of the HIV synthetic peptides. PBL from a Walter Reed Stage 1 patient (1,463 CD4<sup>+</sup> cells mm<sup>-3</sup>) (Fig. 1b) responded as well to FLU as those from the HIV<sup>-</sup> control, and also generated strong IL-2 responses to T1 and TH4.1, with lower but positive responses to T2 and P18.

Received 20 March; accepted 24 April 1989.

1. Watkins, J. C. & Olverman, H. J. *Trends Neurosci.* **10**, 265–272 (1987).
2. Mayer, M. L. & Westbrook, G. L. *Prog. Neurobiol.* **28**, 197–276 (1987).
3. Bowman, C. L. & Kimelberg, H. K. *Nature* **311**, 656–659 (1984).
4. Kettenman, H., Backus, K. H. & Schachner, M. *Neurosci. Lett.* **52**, 25–29 (1984).
5. Gallo, V., Giovannini, C., Suergiu, R. & Levi, G. *J. Neurochem.* **52**, 1–9 (1989).
6. Sontheimer, H., Kettenman, H., Backus, K. H. & Schachner, M. *Glia* **1**, 328–336 (1988).
7. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pfugers Arch. ges. Physiol.* **391**, 85–100 (1981).



TABLE 1 T-helper cell responses of PBL from HIV seropositive and seronegative donors to influenza A virus and four HIV synthetic peptides

| Donor number | IL-2 production response |      |      |       |     | Proliferation response |     |     |       |     | Donor number | IL-2 production response |     |     |       |     | Proliferation response |     |     |       |     |
|--------------|--------------------------|------|------|-------|-----|------------------------|-----|-----|-------|-----|--------------|--------------------------|-----|-----|-------|-----|------------------------|-----|-----|-------|-----|
|              | FLU                      | T1   | T2   | TH4.1 | P18 | FLU                    | T1  | T2  | TH4.1 | P18 |              | FLU                      | T1  | T2  | TH4.1 | P18 | FLU                    | T1  | T2  | TH4.1 | P18 |
| 363*         | 6.7                      | 9.5  | 2.8  | 2.3   | 3.1 | 2.7                    | 1.6 | 2.9 | 0.3   | 0.4 | 516†         | 1.8                      | 1.5 | 1.4 | 1.2   | 1.9 | 1.4                    | 0.6 | 1.4 | 0.1   | 0.1 |
| 471          | 2.6                      | 2.2  | 2.8  | 1.6   | 3.7 | 0.4                    | 0.2 | 0.8 | 0.2   | 0.2 | 575          | 0.9                      | 0.8 | 0.5 | 0.9   | 0.8 | 0.9                    | 1.7 | 2.2 | 0.1   | 0.1 |
| 136          | 2.3                      | 2.5  | 0.6  | 0.3   | 1.1 | 1.1                    | 0.8 | 0.4 | 0.4   | 0.8 | 176          | 1.3                      | 1.4 | 1.1 | 1.3   | 2.0 | 2.9                    | 1.2 | 0.2 | 0.5   | 0.2 |
| 300          | 2.1                      | 0.8  | 0.8  | 1.2   | 1.0 | 1.0                    | 0.3 | 0.3 | 0.2   | 0.3 | 415          | 1.9                      | 1.4 | 0.9 | 0.7   | 1.0 | 1.2                    | 1.3 | 0.6 | 0.8   | 1.0 |
| 346          | 2.1                      | 1.9  | 1.8  | 2.2   | 2.1 | 3.5                    | 5.2 | 2.4 | 2.2   | 2.3 | 756          | 1.0                      | 0.8 | 0.8 | 0.8   | 0.9 | 0.7                    | 1.3 | 1.2 | 0.2   | 0.3 |
| 429          | 6.4                      | 2.1  | 2.1  | 1.6   | 1.0 | 6.4                    | 2.5 | 1.2 | 0.8   | 1.1 | 328          | 1.8                      | 1.0 | 1.6 |       |     | 0.9                    | 1.4 | 1.0 |       |     |
| 149          | 6.2                      | 9.6  | 1.1  |       |     | 1.5                    | 1.5 | 1.1 |       |     | 20           | 1.2                      | 1.6 | 0.6 |       |     | 0.8                    | 1.2 | 1.2 |       |     |
| 146          | 4.7                      | 2.7  | 3.0  |       |     | 0.9                    | 0.5 | 0.6 |       |     | 769          | 1.6                      | 0.6 | 1.0 | 0.9   | 1.0 |                        |     |     |       |     |
| 318          | 12.0                     | 0.9  | 1.0  |       |     | 0.9                    | 0.8 | 0.7 |       |     | 352          | 1.0                      | 0.4 | 0.5 | 0.8   | 0.9 |                        |     |     |       |     |
| 320          | 5.5                      | 1.1  | 0.8  |       |     | 1.7                    | 0.4 | 0.5 |       |     | 373          | 1.7                      | 0.8 | 0.3 | 1.0   | 0.9 |                        |     |     |       |     |
| 326          | 3.7                      | 3.2  | 0.8  |       |     | 3.5                    | 0.7 | 1.4 |       |     |              |                          |     |     |       |     |                        |     |     |       |     |
| 382          | 19.1                     | 20.0 | 24.1 |       |     | 0.9                    | 0.8 | 0.9 |       |     | 43‡          | 9.8                      | 0.4 | 0.3 | 0.8   | 0.3 | 23.7                   | 0.5 | 0.3 | 0.2   | 0.2 |
| 331          | 6.2                      | 1.4  | 2.2  |       |     | 1.1                    | 0.7 | 0.8 |       |     | 44           | 12.1                     | 0.8 | 0.9 | 0.1   | 0.2 | 6.8                    | 0.5 | 0.6 | 0.7   | 0.7 |
| 108          | 2.3                      | 2.1  | 1.5  |       |     | 0.6                    | 2.0 | 2.3 |       |     | 45           | 10.9                     | 1.3 | 1.2 | 0.9   | 0.8 | 8.7                    | 0.9 | 1.3 | 0.6   | 1.0 |
| 186          | 2.2                      | 1.7  | 2.7  |       |     | 0.8                    | 1.1 | 0.8 |       |     | 46           | 5.7                      | 1.0 | 0.9 | 1.1   | 1.0 | 4.1                    | 0.8 | 0.4 | 0.4   | 0.8 |
| 250          | 4.9                      | 5.5  | 2.8  |       |     | 2.4                    | 1.9 | 1.5 |       |     | 47           | 5.7                      | 0.7 | 0.8 | 1.0   | 1.1 | 55.9                   | 0.5 | 0.5 | 0.4   | 0.2 |
| 425          | 3.1                      | 2.0  | 2.3  |       |     | 1.1                    | 2.2 | 1.8 |       |     | 48           | 4.6                      | 1.6 | 0.3 | 0.3   | 0.2 | 42.3                   | 0.8 | 0.6 | 0.2   | 0.1 |
| 253          | 96.8                     | 2.1  | 2.7  | 4.5   | 7.8 |                        |     |     |       |     | 49           | 3.4                      | 1.1 | 1.1 | 0.7   | 0.6 | 3.7                    | 0.1 | 0.2 | 0.3   | 0.2 |
| 772          | 2.8                      | 2.5  | 2.1  | 2.9   | 1.6 |                        |     |     |       |     | 50           | 3.1                      | 1.1 | 0.2 | 0.5   | 0.7 | 15.2                   | 1.1 | 0.9 | 0.8   | 1.0 |
| 399          | 2.9                      | 1.4  | 3.4  | 2.8   | 2.4 |                        |     |     |       |     | 51           | 2.9                      | 0.4 | 0.5 | 1.0   | 0.8 | 19.9                   | 0.7 | 0.6 | 0.8   | 0.2 |
| LW           | 4.5                      | 2.8  | 0.7  | 2.2   | 1.2 |                        |     |     |       |     | 52           | 4.3                      | 0.4 | 0.3 | 0.2   | 0.2 | 7.7                    | 0.8 | 0.5 | 0.9   | 0.5 |
|              |                          |      |      |       |     |                        |     |     |       |     | 53           | 4.4                      | 0.8 | 1.1 | 0.2   | 0.3 | 17.1                   | 1.1 | 1.0 | 0.4   | 0.3 |
| 513†         | 1.1                      | 0.7  | 0.9  | 1.1   | 1.2 | 1.1                    | 0.3 | 0.8 | 0.1   | 0.1 | 54           | 7.8                      | 0.1 | 0.1 | 1.1   | 0.3 | 5.9                    | 0.1 | 0.1 | 0.1   | 0.2 |
| 308          | 0.9                      | 0.6  | 0.7  | 1.1   | 0.6 | 1.3                    | 0.5 | 0.5 | 0.4   | 0.8 | 55           | 7.9                      | 0.3 | 0.2 | 0.6   | 0.2 | 5.9                    | 0.1 | 0.1 | 0.1   | 0.2 |
| 353          | 1.8                      | 0.9  | 1.1  | 0.8   | 1.1 | 1.2                    | 1.2 | 1.1 | 1.3   | 0.4 | 56           | 6.4                      | 1.0 | 1.2 | 0.8   | 0.6 | 6.8                    | 0.6 | 0.7 | 0.8   | 0.6 |
| 757          | 1.1                      | 1.4  | 1.4  | 1.7   | 1.2 | 1.4                    | 0.4 | 0.6 | 1.7   | 1.8 | 57           | 5.8                      | 1.0 | 0.7 | 1.3   | 0.4 | 7.3                    | 0.4 | 0.3 | 0.3   | 0.2 |
| 1            | 1.9                      | 0.9  | 1.1  | 1.0   | 0.6 | 0.6                    | 0.9 | 1.4 | 0.5   | 1.1 | 58           | 10.5                     | 1.0 | 1.2 | 0.3   | 0.3 | 7.6                    | 0.4 | 0.4 |       |     |
| 304          | 1.4                      | 1.0  | 0.8  | 1.2   | 0.9 | 0.7                    | 1.5 | 1.0 | 0.4   | 1.1 | 59           | 5.6                      | 1.2 | 1.0 | 1.1   | 0.4 | 17.5                   | 1.2 | 1.3 |       |     |
| 671          | 1.6                      | 1.0  | 1.2  | 2.0   | 1.7 | 1.3                    | 0.9 | 1.3 | 0.7   | 1.3 | 60           | 12.6                     | 0.3 | 0.6 | 0.1   | 0.4 | 21.4                   | 2.3 | 2.1 |       |     |
| 417          | 1.6                      | 1.5  | 1.0  | 1.0   | 0.7 | 1.1                    | 1.0 | 1.2 | 1.0   | 1.4 | 61           | 11.7                     | 0.7 | 1.2 | 1.1   |     | 15.7                   | 1.1 | 0.9 |       |     |
| 267          | 1.7                      | 1.6  | 0.9  | 0.6   | 0.5 | 1.2                    | 1.1 | 1.3 | 0.4   | 0.3 | 62           | 5.3                      | 0.4 | 0.6 | 1.0   | 1.1 |                        |     |     |       |     |
| 79           | 1.6                      | 0.8  | 0.8  | 0.8   | 1.4 | 1.1                    | 0.8 | 1.5 | 0.1   | 0.1 | 63           | 3.3                      | 0.7 | 0.4 | 0.7   | 0.7 |                        |     |     |       |     |
| 327          | 1.9                      | 1.3  | 1.2  | 1.7   | 1.6 | 2.5                    | 1.5 | 1.4 | 0.1   | 0.1 |              |                          |     |     |       |     |                        |     |     |       |     |

Values indicate stimulation indices (see Fig 1). Underlined values indicate stimulation indices  $\geq 2.0$  for responses to the T1, T2, TH4.1 and P18 peptides. Peptides were tested at 2.5  $\mu$ M and FLU as in Fig. 1. For IL-2 production, all supernatants were tested at 5 serial dilutions in quadruplicate. Although means for 1:4 dilution are shown, they represent results of 20 assays wells. Proliferation assays (day 7) were done in triplicate wells ( $3 \times 10^5$  PBL per well).

\* The first 21 donors were HIV<sup>+</sup> individuals whose PBL generated IL-2 in response to FLU. The donor designated LW is a laboratory worker who was accidentally infected with HIV-1<sub>IIIIB</sub>.

† The second 21 donors were HIV<sup>+</sup> individuals whose PBL did not generate IL-2 in responses to FLU.

‡ Donors 43-63 were HIV<sup>-</sup> individuals whose PBL were used as controls.

PBL from the other Walter Reed Stage 1 patient (583 CD4<sup>+</sup> cells mm<sup>-3</sup>) (Fig. 1c) failed to respond either to FLU or to any of the four peptides. Thus, we have identified two types of HIV<sup>+</sup> individuals, one who responds to FLU and to the four HIV synthetic peptides, and another who does not respond to FLU or to any of the four peptides tested. Using the experimental design shown in Fig. 1, we tested PBL from an additional 40 HIV<sup>+</sup> individuals and 20 HIV<sup>-</sup> control donors (Table 1). We also tested all donors for response to tetanus toxoid and found a complete concordance with their responsiveness to FLU.

Of the 21 HIV<sup>+</sup> donors who responded to FLU by IL-2 production, only 5 of the 17 tested responded with T-cell proliferation against FLU. In contrast, all of the 19 normal controls responded by T-cell proliferation against FLU. This result indicates that capacity for T-cell proliferation is lost earlier in HIV infection than for IL-2 production. This observation may in part account for earlier failures to detect responses to HIV antigens<sup>10</sup>. To be sure that the failure to observe proliferation despite a positive IL-2 response was not simply due to a difference in the kinetics of the T-cell proliferative response between the HIV<sup>+</sup> individuals and the controls, we investigated the kinetics in three HIV<sup>+</sup> individuals and one control. The control was positive for IL-2 production on day 7, for proliferation on days 5, 7, and 9 and weakly responsive on day 11 of culture. By contrast, the three HIV<sup>+</sup> donors were all negative for proliferative response on days 5, 7, 9, and 11, despite being reproducibly positive for IL-2 production in the same experiment (data not shown). Also, five donors who were unresponsive to FLU by both assays remained negative at all four time points. Therefore, the difference is a qualitative defect in proliferative response and not just a difference in kinetics. Follow up studies of these

donors may indicate whether the difference observed at this early time point after infection is of prognostic significance.

Among the 21 HIV<sup>+</sup> donors who responded to FLU, 14 responded to T1 and 12 responded to T2. PBL from 6 of 10 of these same donors tested with the TH4.1 peptide responded, and 5 of 10 tested with peptide P18 responded. Only two of the 21 HIV<sup>+</sup> donors who were unresponsive to FLU responded to any of the four peptides. None of the 21 HIV<sup>-</sup>/FLU<sup>+</sup> control donors responded to any of the peptides. Two of the HIV<sup>+</sup> donors who responded to FLU and to one or more of the HIV

TABLE 2 Lack of correlation between HLA class II antigen expression in HIV<sup>+</sup>/FLU<sup>+</sup> donors and IL-2 response to HIV synthetic peptides

| Donor | HLA antigens |      | IL-2 response to peptide |    |       |     |
|-------|--------------|------|--------------------------|----|-------|-----|
|       | DR           | DQ   | T1                       | T2 | TH4.1 | P18 |
| 363   | 13, 52       | 1    | +                        | +  | +     | +   |
| 471   | 2, 8, 52     | 1    | +                        | +  | —     | +   |
| 149   | 4, 52, 53    | 3    | +                        | —  | NT    | NT  |
| 146   | 3, 7         | 2, 3 | +                        | +  | NT    | NT  |
| 318   | 3, 52        | 1    | —                        | —  | NT    | NT  |
| 320   | 3, 4, 52     | ?    | —                        | —  | NT    | NT  |
| 326   | 1, 8         | 1    | +                        | —  | NT    | NT  |
| 382   | 2, 3, 52     | ?    | +                        | +  | NT    | NT  |
| 331   | 11, 13, 52   | 1, 3 | —                        | +  | NT    | NT  |
| 108   | 1, 9, 52, 53 | 1, 3 | +                        | —  | NT    | NT  |
| 425   | 1, 4, 52, 53 | 1    | +                        | +  | NT    | NT  |
| 772   | 8, 13, 52    | 1    | +                        | +  | +     | —   |
| LW    | 1, 2, 52     | 1    | +                        | —  | +     | —   |

NT, not tested.

peptides by IL-2 production were available for re-testing one to three months later, and were found to be reproducibly positive against the same antigens. One HIV<sup>+</sup> donor who showed no response to FLU or to any of the peptides remained negative to all of the antigens when retested one year later. PBL from 13 of the 21 HIV<sup>+</sup> donors who responded to FLU were typed for HLA. In this small sample, no correlation was noted between the expression of a particular HLA class I or class II antigen and ability to respond to any of the synthetic peptides (Table 2).

The individuals who responded to the synthetic peptides were two-to-three times as frequent among HIV<sup>+</sup>/FLU<sup>+</sup> donors as among all HIV<sup>+</sup> donors (Table 3). All donors, irrespective of their HIV status, were responsive to HLA alloantigens, which is indicative of their ability to respond to some antigenic stimulus by IL-2 production. It has recently been demonstrated that T-helper cell responses by human PBL to FLU require MHC self-restricted CD4<sup>+</sup> T-helper cells whereas the response to alloantigens can use either the CD4<sup>+</sup> or CD8<sup>+</sup> pathway of T-helper cell activity<sup>7</sup> (C.S. Via, G. Tsokos and G.M.S., manuscript in preparation). Also, 40–50% of asymptomatic HIV<sup>+</sup> individuals and AIDS patients have a selective defect in CD4-mediated but not in CD8-mediated T-helper cell function (M. C. *et al.*, submitted). Therefore, almost all individuals responsive to one or more of the HIV synthetic peptides also responded to FLU, probably because this group of 21 of the 42 HIV<sup>+</sup> donors tested retained intact T-helper cell function to a CD4-dependent antigen. In contrast, 20 of the 21 patients unresponsive to FLU may also have failed to respond to the peptides because they had lost responsiveness to any CD4-dependent antigens. Thus, it is important when testing any HIV<sup>+</sup> individuals

TABLE 3 Positive T-helper cell responses to the four HIV synthetic peptides as a function of ability to respond to influenza A virus

| Donor category                      | IL-2 production |                |               |               | Proliferation |               |              |              |
|-------------------------------------|-----------------|----------------|---------------|---------------|---------------|---------------|--------------|--------------|
|                                     | T1              | T2             | TH4.1         | P18           | T1            | T2            | TH4.1        | P18          |
| All HIV <sup>+</sup>                | 14/42<br>(33%)  | 12/42<br>(29%) | 6/29<br>(21%) | 6/29<br>(21%) | 4/35<br>(11%) | 4/35<br>(11%) | 1/22<br>(5%) | 1/22<br>(5%) |
| HIV <sup>+</sup> , FLU <sup>+</sup> | 14/21<br>(67%)  | 12/21<br>(57%) | 6/10<br>(60%) | 5/10<br>(50%) | 4/17<br>(24%) | 3/17<br>(18%) | 1/6<br>(17%) | 1/6<br>(17%) |

Fractions and per cent (in parentheses) given for each donor category. All of the HIV seronegative control donors responded to FLU, but none of them responded to any of the four HIV synthetic peptides.

for T-helper cell responses to HIV antigens to establish whether these patients have retained an intact CD4 T-helper cell pathway.

Of the 21 FLU-responsive HIV<sup>+</sup> donors, 18 responded to at least one of the four T-cell epitopes. Of the 10 donors we were able to test with all four peptides, nine responded to at least one peptide. Therefore, these four helper T-cell epitopes are sufficient to elicit responses in 85–90% of an HLA-diverse group of patients. For the purposes of vaccine development, it has been feared that many T-cell epitopes would be necessary to cover most or all of the outbred human population with extensive HLA polymorphism. But the present results are much encouraging: only a few selected epitopes may be sufficient. □

Received 18 November 1988; accepted 19 April 1989.

1. Cease, K. B. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 4249–4253 (1987).
2. Berzofsky, J. A. *et al.* *Nature* **334**, 706–708 (1988).
3. DeLisi, C. & Berzofsky, J. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7048–7052 (1985).
4. Spouge, J. L. *et al.* *J. Immunol.* **138**, 204–212 (1987).
5. Margalit, H. *et al.* *J. Immunol.* **138**, 2213–2229 (1987).
6. Berzofsky, J. A. *et al.* *Immunol. Rev.* **98**, 9–52 (1987).
7. Shearer, G. M. *et al.* *J. Immunol.* **137**, 2514–2521 (1986).
8. Lane, H. C. *et al.* *New Engl. J. Med.* **313**, 79–84 (1985).
9. Redfield, R. R. *New Engl. J. Med.* **314**, 131–132 (1986).
10. Warren, B. *et al.* *J. Virol.* **61**, 2017–2023 (1987).
11. Takahashi, H. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 3105–3109 (1988).

ACKNOWLEDGEMENTS. We thank Drs Steven M. Barks and David Alling (NIAID, NIH) for statistical analysis of the data for possible HLA associations, and Ms Judy Kress for preparing the manuscript. The opinions expressed herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Air Force or the Department of Defense.

## An engineered poliovirus chimaera elicits broadly reactive HIV-1 neutralizing antibodies

David J. Evans\*, Jane McKeating†, Janet M. Meredith\*, Karen L. Burke\*, Kersi Katrak‡, Ann John‡, Morag Ferguson‡, Philip D. Minor‡, Robin A. Weiss† & Jeffrey W. Almond\*

\* Department of Microbiology, University of Reading, London Road, Reading RG1 5AQ, UK

† Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, UK

‡ National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3QR, UK

THE Sabin type 1 vaccine strain of poliovirus is probably the safest and most successful live-attenuated vaccine virus used in humans. Its widespread use since the early 1960s has contributed significantly to the virtual eradication of poliomyelitis in developed countries. We have reported previously the construction of an intertypic antigen chimaera of poliovirus, based on the Sabin 1 strain, and proposed that this virus could be modified to express on its surface antigenic determinants from other pathogens<sup>1</sup>. We describe here the construction and characterization of a poliovirus

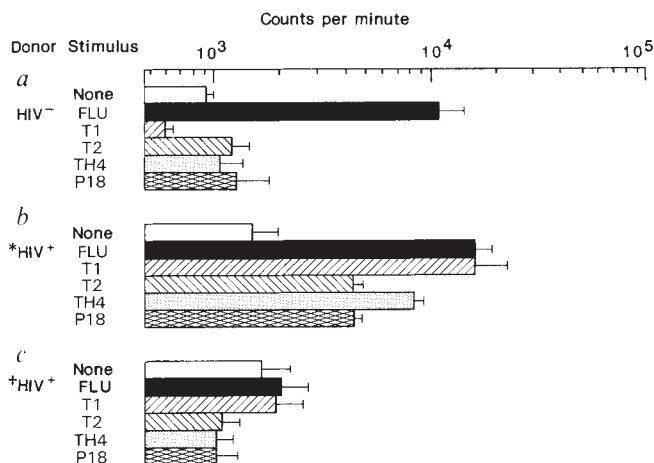


FIG. 1 IL-2 production by PBL from an HIV seronegative (a) and two HIV seropositive Walter Reed Stage 1 patients (b and c).

METHOD. PBL were unstimulated, or stimulated with influenza A virus (A/Hong Kong RX73, H3N2, grown in chicken eggs, final culture dilution 1:1,000) or with the HIV synthetic peptides T1, T2, TH4 or P18 at 2.5 µM. EnvT1 and T2 correspond to amino-acid residues 428–443 and 112–124 of gp120 (IIIB isolate) respectively<sup>1</sup>. Peptide TH4.1 corresponds to residues 834–848 of gp160 (P. Hale *et al.*, submitted). P18, corresponding to residues 315–329 of gp160 (IIIB), is a major epitope for murine anti-HIV cytolytic effector cells<sup>11</sup>. PBL ( $3 \times 10^6$ ) were cultured in 2 ml RPMI-1640 medium supplemented with 5% pooled AB<sup>+</sup> human serum for 7 d in the presence of anti-TAC (IL-2 receptor p55 chain) monoclonal antibody (to prevent IL-2 consumption). Supernatants of these cultures were collected, and five twofold dilutions of the supernatants were added to cultures of the IL-2-dependent CTLL cell line. Twenty-four hours later, the stimulated CTLL cultures were pulsed with [<sup>3</sup>H]thymidine; thymidine incorporation was determined 18 h later, and is expressed in counts per minute. The data points shown in Fig. 1 and Table 1 are for a culture supernatant dilution of 1:4. We have used a stimulation index (ratio of counts per minute in stimulated cultures to that in unstimulated cultures) of greater than 2.0 as an indication of positive response. For key, see Table 1.

antigen chimaera containing an epitope from the transmembrane glycoprotein (gp41) of human immunodeficiency virus type 1 (HIV-1). In antibody absorption experiments, the virus chimaera inhibited neutralization of HIV-1 by anti-peptide monoclonal antibodies specific for the gp41 epitope and significantly reduced the group specific neutralizing activity of HIV-1-positive human sera. Rabbit antisera raised by subcutaneous injection of the polio/HIV chimaera in adjuvant was shown to be specific for HIV-1 gp41 in peptide-binding assays and by western blotting. Moreover, the antisera neutralized a wide range of American and African HIV-1 isolates and also inhibited virus-induced cell fusion. Monoclonal antibodies against the HIV-1 derived regions of the chimaera also neutralized HIV-1. These results establish the potential of using poliovirus for the presentation of foreign antigens and suggest that Sabin 1 poliovirus/HIV chimaeras could offer an approach to the development of an HIV vaccine.

We have used recombinant DNA techniques to introduce antigenic domains from the capsid proteins of poliovirus type 3 into poliovirus type 1 to produce a virus chimaera of dual antigenicity and immunogenicity<sup>1</sup>. These and related observations<sup>2,3</sup> encouraged us to develop the Sabin type 1 vaccine strain of poliovirus as an expression vector/vaccine for the presentation of antigenic structures from unrelated pathogens. To facilitate the production of virus antigen chimaeras, we have constructed a cassette vector (pCAS1) by engineering unique restriction sites flanking the region encoding amino acids 91–102 of capsid protein VP1 (Fig. 1a) in a full-length complementary DNA of Sabin 1 cloned in a pBR322-derived vector. This region of VP1 is known to elicit neutralizing antibodies<sup>4</sup>, and on the three-dimensional structure of poliovirus is seen to form a distinct surface projection at the pentameric apex of the icosahedral particle<sup>5</sup>. Use of pCAS1 allows the replacement of the cDNA corresponding to antigenic site 1 by synthetic oligonucleotides encoding an antigenic determinant of choice.

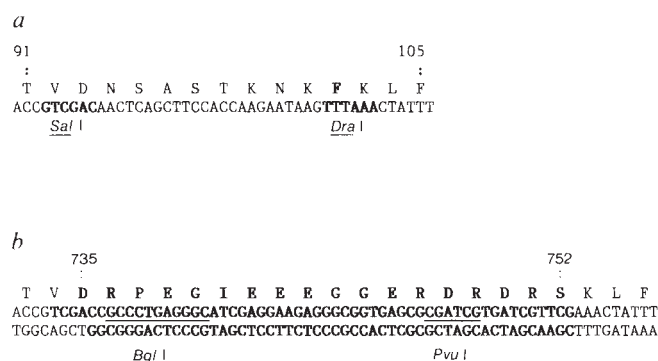


FIG. 1 Construction of the poliovirus: HIV-1 antigen chimaera S1/env/3. **a**, The nucleotide and amino-acid sequence of antigenic site 1 of the Sabin 1-derived cassette vector pCAS1. The highlighted residue represents a single amino-acid change (Asp to Phe at position 102 of VP1) generated as a result of the introduction of the *Dra I* site. **b**, The nucleotide and amino acid sequence of the corresponding region in the antigen chimaera S1/env/3. Residues in bold type indicate the 18 amino acids derived from HTLV-III<sub>B</sub> gp41 (amino acids 735–752) (ref. 25).

**METHODS.** **a**, Unique *Sal I* and *Dra I* restriction sites were engineered into a full-length cDNA of Sabin 1, preceded by a synthetic T7 promoter, and cloned in the pBR322-derived vector pFB1 (Pharmacia). The precise construction details of the Sabin 1-based expression vector (pCAS1) will be reported elsewhere. **b**, Complementary oligonucleotides were annealed and ligated with *Sal I*/*Dra I* digested pCAS1, and recombinants screened for *Bgl I* and *Pvu I* restriction sites (underlined). A recombinant pS1/env/3 was linearized with *Nae I*, which cuts within vector sequences of the construct, and used as a template in a T7 transcription reaction<sup>26</sup> prior to transfection of sub-confluent Hep2C monolayers. A cytopathic effect was observed after 4 days and the RNA sequence of ~200 base pairs spanning antigenic site 1 (2,650–2,850) of the recovered chimaeric virus was confirmed by sequencing<sup>27</sup>.

The envelope glycoprotein (env) of HIV-1 displays extensive sequence heterogeneity and antigenic variation<sup>6,7</sup>. Antigenically conserved sites within env must exist, however, as it is possible to show neutralization of a range of HIV-1 isolates with immune sera from infected individuals<sup>8</sup>, or with antisera raised to defined regions of env<sup>9</sup>. Prospective vaccines should ideally contain such conserved sites so that they elicit antibodies capable of reacting with a wide range of virus isolates. We have chosen a well characterized epitope from the transmembrane glycoprotein gp41 of HIV-1 (amino acids 735–752). A synthetic peptide of this region is recognized in enzyme-linked immunosorbent assay by antisera from seropositive patients and induces antibodies which neutralize a range of HIV-1 isolates<sup>10–12</sup>, indicating that it is a group-specific epitope. The construction of a chimaeric poliovirus containing this sequence is detailed in Fig. 1.

The recovered virus, designated S1/env/3, was neutralized and recognized in antigen-blocking tests<sup>13</sup> by polyclonal poliovirus type 1 antisera, and by monoclonal antibodies (mAbs) directed against antigenic sites 2 and 3 of the Sabin 1 strain, but not by antibodies specific for antigenic site 1.

The recent observation that the structure of antigenic site 1 may influence poliovirus host range<sup>2,14</sup> prompted us to investigate the interaction of S1/env/3 with the poliovirus receptor. S1/env/3 infection of Hep2C (human epithelial carcinoma cells) monolayers was blocked by a mAb specific for the receptor<sup>15</sup>, thus demonstrating that the chimaera was still dependent on the normal poliovirus cellular receptor for attachment and entry into cells.

Preincubation with purified S1/env/3 but not Sabin 1 abrogated the HTLV-III<sub>B</sub> neutralizing activity of two anti-peptide mAbs, ED6 and LA9, and reduced the neutralization of HTLV-III<sub>B</sub> by five of six human sera. Neutralization of HTLV-III<sub>B</sub> by a control antibody (110.3) specific for gp120 amino acids 307–321 (refs 16, 17), was not inhibited by S1/env/3 or Sabin 1.

The immunogenic potential of S1/env/3 was investigated in rabbits, and the neutralizing activity against HIV-1 was determined by infectivity inhibition and plaque reduction assays

TABLE 1 S1/env/3 inhibition of HTLV-III<sub>B</sub> neutralization

|                              | Residual HTLV-III <sub>B</sub> neutralization titre after incubation with |         |          |
|------------------------------|---------------------------------------------------------------------------|---------|----------|
|                              | Mock                                                                      | Sabin 1 | S1/env/3 |
| <b>Monoclonal antibodies</b> |                                                                           |         |          |
| ED6                          | 640                                                                       | 640     | 0        |
| LA9                          | 640                                                                       | 640     | 0        |
| 110.3                        | 1,000                                                                     | 1,000   | 1,000    |
| <b>Human sera</b>            |                                                                           |         |          |
| 1                            | 160                                                                       | 160     | 20       |
| 2                            | 40                                                                        | 40      | 10       |
| 3                            | 40                                                                        | 40      | 10       |
| 4                            | 80                                                                        | 80      | 20       |
| 5                            | 320                                                                       | 160     | 40       |
| 6                            | 80                                                                        | 80      | 80       |

Results are expressed as the reciprocal of the serum dilution giving >90% reduction in HIV titre<sup>8</sup> following pre-incubation of the mAbs or human immune sera with culture medium (Mock), Sabin 1 or S1/env/3. LA9 and ED6 are IgM mAbs specific for the gp41 epitope (HTLV-III<sub>B</sub> amino acids 735–752). 110.3 is an IgG1 mAb specific for the LAV-1 type specific neutralization loop in the second conserved domain of gp120 (amino acids 307–321) (refs 16, 17). Human immune sera numbered 1–6 represent anonymous British HIV<sup>+</sup> blood donors. Sucrose-purified S1/env/3 or Sabin 1 ( $5 \times 10^4$  TCID<sub>50</sub> (50% tissue culture infectious dose) units) were preincubated with a 1:10 dilution of ED6, LA9, a 1:100 dilution of 110.3 and a 1:1 dilution of each human serum for 1 h at 37 °C. Residual HIV-neutralizing activity was determined by incubating dilutions of the antibody/virus mixture with  $10^3$  infectious units (TCID<sub>50</sub>) of HTLV-III<sub>B</sub> for 1 h at 37 °C. Aliquots (100 µl) of medium containing  $2 \times 10^4$  C8166 cells were added, and the presence of syncytia recorded after 48 h as an indication of HIV infection.



(Table 2). All rabbit antisera showed neutralizing activity against HTLV-III<sub>B</sub>, confirming that the chimaera has immunogenic potential. Antiserum R1, which contained the highest anti-poliovirus activity, was further tested against a range of HIV-1 isolates. This antiserum neutralized the entire test panel at various titres, including the African isolates CBL4 (Tanzania) and Z84 (Zaire). The titres observed in both neutralization assays used were in good agreement and were comparable to those of human sera (see Table 1). Antisera R7 and R8 also neutralized the HIV-1 isolates tested (HTLV-III<sub>RF</sub> or Z84), although the titres observed were lower against both HIV-1 and S1/env/3. The HIV-1-neutralizing activity of antiserum R1 was absorbed out by pre-incubation with S1/env/3, but not with Sabin 1, confirming that the activity was induced by the gp41 epitope and not by a chance cross-reactive poliovirus epitope (data not shown). Early syncytium formation in a mixture of HTLV-III<sub>B</sub> producing cells and uninfected C8166 cells (CD4<sup>+</sup>, HTLV-I-transformed T cells)<sup>18</sup> was inhibited by antiserum R1, although at a lower titre than that determined for virus inhibition (data not shown).

Four mAbs specific for S1/env/3 were characterized in terms of their reaction with HIV-1. One mAb (1577) displayed neutralizing activity against all HIV-1 isolates tested, including the three African strains CBL4, Z84 and Z129 (Table 2). Monoclonal antibodies 1575 and 1583 neutralized only some of the isolates, suggesting that they recognize a defined epitope within the gp41 735–752 region, which is less well conserved. Monoclonal antibody 1578, displayed no HIV-1-neutralizing activity, suggesting that it recognized an epitope formed from both HIV-1 and poliovirus amino acids.

The specificity of the HIV-1 immune response to S1/env/3 was demonstrated by western blotting and by a peptide binding assay (Fig. 2). Antiserum R1 reacted with the envelope glycoprotein precursor gp160 and gp41 in western blots, whereas pre-immune sera were negative (Fig. 2a). All the rabbit antisera bound specifically to a linear synthetic peptide corresponding to the HTLV-III<sub>B</sub> epitope present on the S1/env/3 chimaera (Fig. 2b), whereas they failed to bind a 15-amino-acid peptide derived from the type-specific neutralization epitope on gp120 (residues 307–321). Pre-immune sera and control rabbit hyper-immune Sabin 1 antisera displayed no specific binding to either the gp41 or gp120 derived peptides (Fig. 2c). The three S1/env/3 monoclonal antibodies that neutralized HIV-1 (1575, 1577, 1583) also reacted with the gp41 735–752 peptide in binding assays (data not shown).

The poliovirus chimaera described here is based on the safe and effective Sabin type 1 live attenuated vaccine strain. The substitution of eight poliovirus residues by eighteen HIV-1 (HTLV-III<sub>B</sub>) specified amino acids gave rise to a viable virus which possessed some of the antigenic properties of HIV-1. This substitution did not involve the alteration of amino acids believed to be important to the attenuation phenotype of Sabin 1, so we expect the virus chimaera to have retained this desirable property. The results illustrate the flexibility of the surface of the poliovirus particle and indicate that it may be possible to generate further chimaeras containing, for example, additional HIV-1, HIV-2, or composite B- and T-cell epitopes<sup>19</sup>. The gp41 sequence chosen here induced HIV-1 group-specific neutralizing antibodies in rabbits. The inhibition, by S1/env/3, of HTLV-III<sub>B</sub> neutralizing activity in human immune sera suggests that, in five out of six individuals, a significant proportion of the cross-reactive immune response may be directed against this epitope. The antisera also inhibited the fusion of HTLV-III<sub>B</sub> infected and uninfected C8166 cells, suggesting that fusion inhibition may be a characteristic of antibodies against this epitope.

It is possible that different regions of the envelope glycoproteins of HIV may elicit antibodies which play different roles in immunopathogenesis and protection, for example, antibody-dependent cell cytotoxicity, Fc-mediated virus enhancement, fusion inhibition or attachment inhibition. Vaccinia virus

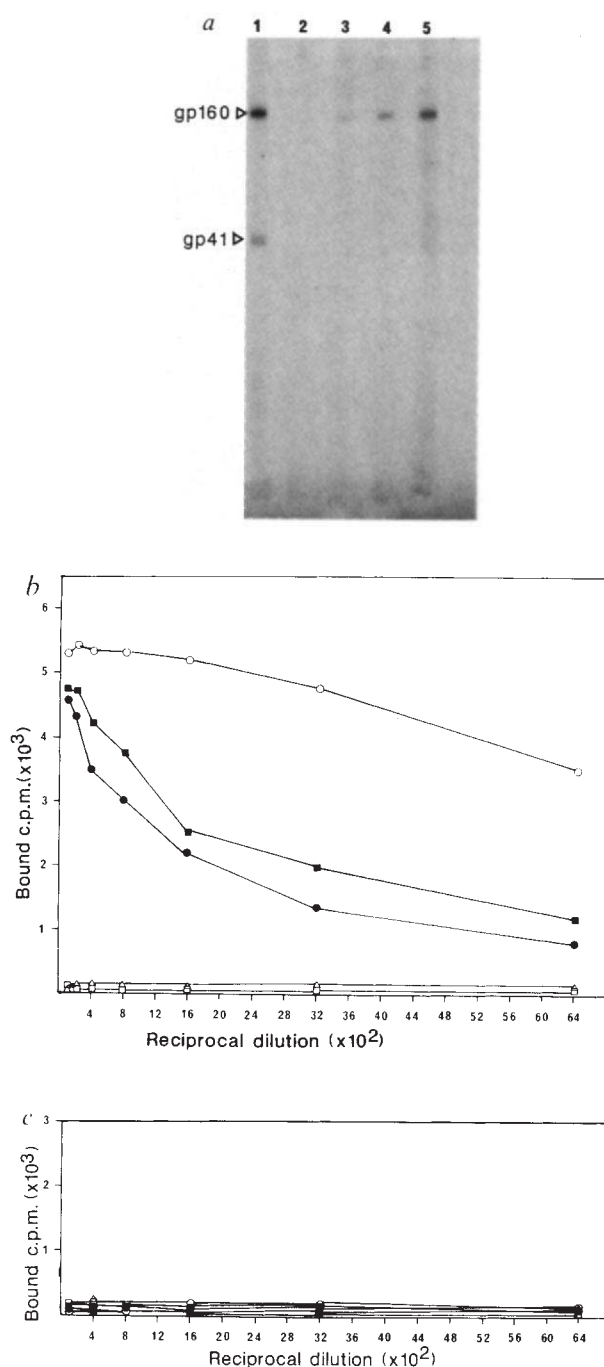


FIG. 2 Specificity of the immune response to S1/env/3. *a*, Analysis of the immune response to S1/env/3 by western blot. Reactivity of a human mAb DZ specific for HIV gp41 (lane 1), pre-immune R1 (lane 2), and S1/env/3 antiserum R1 bleed 1 (lane 3), bleed 2 (lane 4) and bleed 3 (lane 5) to HIV-1 SDS-PAGE-separated viral proteins. Gp160 and gp41 are indicated. *b* and *c*, Binding of S1/env/3 antisera (R1) bleeds 1 (●), 2 (■) and 3 (○), pre-immune serum (□), and Sabin 1 antiserum (Δ), to (*b*) gp41 peptide amino acids 735–752 of HTLV-III<sub>B</sub> and to (*c*) gp120 peptide amino acids 307–321 of HTLV-III<sub>B</sub>.

**METHODS.** *a*, Antisera were incubated with HIV WB strips according to the manufacturers instructions (DuPont), and bound antibodies were detected using [<sup>125</sup>I]Protein A (Amersham) at  $2 \times 10^5$  c.p.m. ml<sup>-1</sup>. *b* and *c*, Flexible 96-well microtitre plates were coated with gp41 or gp120 peptide at  $10 \mu\text{g ml}^{-1}$  (50  $\mu\text{l}$  per well) in carbonate buffer, pH 9.6, for 12 h at 4 °C, washed with PBS/0.05% Tween and incubated with 1% bovine serum albumin (BSA) in PBS/0.05% Tween for 1 h at room temperature. Antiserum (50  $\mu\text{l}$ ) diluted in PBS/1% BSA/0.05% Tween was added and incubated for 1 h at 37 °C. After washing, bound antibodies were detected by incubation with [<sup>125</sup>I]Protein A (Amersham,  $10^5$  c.p.m. ml<sup>-1</sup>, 50  $\mu\text{l}$  per well) for 1 h at 37 °C.

TABLE 2 Neutralization of HIV-1 infectivity and inhibition of syncytium formation by antisera and monoclonal antibodies

| Neutralization by S1/env/3 antisera              |                       |                        |                                 |            |            |            |          |
|--------------------------------------------------|-----------------------|------------------------|---------------------------------|------------|------------|------------|----------|
| Antiserum                                        | HTLV-III <sub>B</sub> | HTLV-III <sub>RF</sub> | Reciprocal neutralization titre |            |            |            | S1/env/3 |
|                                                  |                       |                        | Virus strain                    |            | CBL4       | Z84        |          |
|                                                  |                       |                        | SF2                             | SF33       |            |            |          |
| R1                                               |                       |                        |                                 |            |            |            |          |
| Pre-immune                                       | <10                   | <10                    | <10                             | <10        | <10        | <10        | <10      |
| Final bleed                                      | 80<br>(40)            | 80<br>(40)             | 80<br>(160)                     | 40<br>(40) | 10<br>(80) | 40<br>(40) | >28,960  |
| R7                                               |                       |                        |                                 |            |            |            |          |
| Pre-immune                                       | <10                   | <10                    | NT                              | NT         | NT         | <10        | <10      |
| Final bleed                                      | 20                    | 40                     | NT                              | NT         | NT         | 40         | >2,560   |
| R8                                               |                       |                        |                                 |            |            |            |          |
| Pre-immune                                       | <10                   | <10                    | NT                              | NT         | NT         | NT         | <10      |
| Final bleed                                      | 40                    | 20                     | NT                              | NT         | NT         | NT         | >2,560   |
| R9                                               |                       |                        |                                 |            |            |            |          |
| Pre-immune                                       | <10                   | <10                    | NT                              | NT         | NT         | NT         | <10      |
| Final bleed                                      | 40                    | 10                     | NT                              | NT         | NT         | NT         | 1,280    |
| R19, R20 and R21                                 |                       |                        |                                 |            |            |            |          |
| Pre-immune                                       | <10                   |                        |                                 |            |            |            |          |
| Post-immune                                      | <10                   |                        |                                 |            |            |            |          |
| Neutralization by S1/env/3 monoclonal antibodies |                       |                        |                                 |            |            |            |          |
| mAb                                              | HTLV-III <sub>B</sub> | HTLV-III <sub>RF</sub> | Reciprocal neutralization titre |            |            |            | S1/env/3 |
|                                                  |                       |                        | Virus strain                    |            | Z84        | Z129       |          |
|                                                  |                       |                        | SF33                            | CBL4       |            |            |          |
| 1575                                             | 40                    | 20                     | 40                              | 10         | 10         | <10        | 2,560    |
| 1577                                             | 40                    | 160                    | 80                              | 20         | 20         | 10         | 640      |
| 1578                                             | <10                   | <10                    | <10                             | <10        | <10        | <10        | 640      |
| 1583                                             | 40                    | 40                     | 20                              | <10        | <10        | <10        | >28,960  |

Rabbit R1 was immunized intradermally with 0.1 ml ( $\sim 10^8$  TCID<sub>50</sub> ml<sup>-1</sup>) of sucrose-purified S1/env/3 in complete Freund's adjuvant, and boosted subcutaneously at two-week intervals with the same virus preparation in incomplete Freund's adjuvant. Rabbits R7-R9 (S1/env/3) and R19-R21 (Sabin 1) were immunized intramuscularly with 0.5 ml tissue culture fluid ( $\sim 10^8$  TCID<sub>50</sub> ml<sup>-1</sup>) in complete Freund's adjuvant, and boosted at two-week intervals in a similar manner. Neutralization titres were determined by incubating 10  $\mu$ l of heat-inactivated antiserum with 40  $\mu$ l virus supernatant containing  $10^3$  infectious units of HIV-1 at 37 °C for 1 h. Residual HIV-1 infectivity was measured by the infectivity inhibition assay<sup>8</sup> described in Table 1. Antiserum R1 was also tested for HIV-1 neutralizing activity in a plaque-reduction assay (ref. 24 and J. McK. *et al.*, in preparation) on the sensitive MT4 cell line. The results obtained by this independent assay are shown in brackets. Results are expressed as the reciprocal of the serum dilution giving >90% reduction in HIV infectivity or plaque formation. Also shown is the reciprocal neutralization titre of S1/env/3 antisera with 100 TCID<sub>50</sub> units of the homologous virus. NT, not tested. Murine monoclonal antibodies were raised as described previously<sup>1,3</sup>, and screened against S1/env/3 and Sabin 1 in antigen-blocking tests (data not shown). The S1/env/3 and HIV-1 neutralizing activity of the ascites from four hybridomas was determined. Results are expressed as the reciprocal of the mAb ascitic fluid dilution giving >90% reduction in HIV-1 titre<sup>8</sup>, or neutralizing 100 TCID<sub>50</sub> units of S1/env/3.

constructs expressing env elicit antibodies in humans that neutralize divergent HIV-1 isolates<sup>20</sup>. But the vaccinia-based recombinants<sup>21</sup> used so far express whole or slightly truncated gp160, and thus give rise to a general antibody response. The poliovirus system allows the selective expression of defined regions of HIV proteins in an immunoprominent position, and therefore enables specific immune responses to be induced and analysed. The system also facilitates the generation of monoclonal antibodies against defined HIV sequences which may provide valuable diagnostic reagents.

Poliovirus offers considerable advantages over non-replicating systems such as free peptides or hepatitis B antigen particles<sup>22</sup>, as it induces a secretory as well as a systemic humoral immune response<sup>23</sup>. The presence of secretory antibodies in mucous fluids may reduce the efficiency of transmission of HIV. The immune response elicited by S1/env/3 after oral administration to primates is currently being evaluated. This should provide information on whether poliovirus chimaeras offer a realistic strategy for the development of HIV vaccines. □

- Minor, P. D., Ferguson, M., Evans, D. M. A., Almond, J. W. & Icenogle, J. P. *J. gen. Virol.* **67**, 1283-1291 (1986).
- Hogle, J. M., Chow, M. & Filman, D. J. *Science* **229**, 1358-1365 (1985).
- Modrow, S. *et al. J. Virol.* **61**, 570-578 (1987).
- Starich, B. R. *et al. Cell* **45**, 637-648 (1986).
- Weiss, R. A. *et al. Nature* **324**, 572-575 (1986).
- Ho, D. D., Kaplan, J. C., Rackauskas, I. E. & Gurney, M. E. *Science* **239**, 1021-1023 (1988).
- Kennedy, R. C. *et al. Science* **231**, 1556-1559 (1986).
- Chanh, T. C. *et al. EMBO J.* **5**, 3065-3071 (1986).
- Dalglish, A. G. *et al. Virology* **165**, 209-215 (1988).
- Ferguson, M. *et al. J. gen. Virol.* **65**, 197-201 (1984).
- Murray, M. G. *et al. Science* **241**, 213-215 (1988).
- Minor, P. D., Pipkin, P. A., Hockley, D., Schild, G. C. & Almond, J. W. *Virus Res.* **1**, 203-212 (1984).
- Thomas, E. K. *et al. AIDS* **2**, 25-30 (1988).
- Linsley, P. S., Ledbetter, J. A., Kinney-Thomas, E. & Shiu-Lok, H. *J. Virol.* **62**, 3695-3702 (1988).
- Clapham, P. R. *et al. Virology* **158**, 44-51 (1987).
- Francis, M. J. *et al. Nature* **300**, 168-170 (1987).
- Zagury, D. *et al. Nature* **332**, 728-731 (1988).
- Chakrabarti, S., Robert-Guroff, M., Wong-Staal, F., Gallo, R. C. & Moss, B. *Nature* **320**, 535-537 (1986).
- Michel, M.-L. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 7957-7961 (1988).
- Ogra, P. L. & Ogra, S. S. *J. Immun.* **110**, 1307-1311 (1973).
- Harada, S. *et al. Science* **229**, 563-565 (1985).
- Ratner, L. *et al. Nature* **313**, 277-284 (1986).
- van der Werf, S., Bradley, J., Wimmer, E., Studier, F. W. & Dunn, J. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5080-5084 (1983).
- Rico-Hesse, R., Pallansch, M. A., Nottay, B. K. & Kew, O. M. *Virology* **160**, 311-322 (1987).

Received 28 March; accepted 5 May 1989.

- Burke, K. L., Dunn, G., Ferguson, M., Minor, P. D. & Almond, J. W. *Nature* **332**, 81-82 (1988).
- Martin, A. *et al. EMBO J.* **7**, 2839-2847 (1988).
- Murray, M. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 3203-3207 (1988).

ACKNOWLEDGEMENTS. We thank D. H. Moon for oligonucleotide synthesis, R. C. Gallo, J. A. Levy and D. Zagury for HIV-1 strains, R. Kennedy and J. McClure at Genetic Systems/Oncogen for gp41 and gp120 mAbs, C. Desranges for the human mAb DZ, C. Fricker for rabbit antiserum and J. N. Weber for human sera. This study was supported by the MRC Aids Directed Programme.

# An African primate lentivirus (SIV<sub>sm</sub>) closely related to HIV-2

Vanessa M. Hirsch\*, Robert A. Olmsted\*,  
Michael Murphey-Corb†, Robert H. Purcell‡  
& Philip R. Johnson\*

\* Division of Molecular Virology and Immunology,  
Department of Microbiology, Georgetown University, NIH/Twinbrook II,  
12441 Parklawn Drive, Rockville, Maryland 20852, USA  
† Delta Regional Primate Research Center, Tulane University,  
Covington, Maryland Louisiana 70433, USA  
‡ Laboratory of Infectious Diseases, NIAID, NIH, Bethesda,  
Maryland 20852, USA

THE ancestors of the human immunodeficiency viruses (HIV-1 and HIV-2) may have evolved from a reservoir of African non-human primate lentiviruses, termed simian immunodeficiency viruses (SIV)<sup>1</sup>. None of the SIV strains characterized so far are closely related to HIV-1<sup>2-6</sup>. HIV-2, however, is closely related to SIV (SIV<sub>mac</sub>) isolated from captive rhesus macaques (*Macaca mulatta*)<sup>7</sup>. SIV infection of feral Asian macaques has not been demonstrated by serological surveys<sup>8,9</sup>. Thus, macaques may have acquired SIV in captivity by cross-species transmission from an SIV-infected African primate. Sooty mangabeys (*Cercocebus atys*), an African primate species indigenous to West Africa, however, are infected with SIV (SIV<sub>sm</sub>) both in captivity<sup>9-11</sup> and in the wild (P. Fultz, personal communication). We have molecularly cloned and sequenced SIV<sub>sm</sub> and report here that it is closely related to SIV<sub>mac</sub> and HIV-2. These results suggest that SIV<sub>sm</sub> has infected macaques in captivity and humans in West Africa and evolved as SIV<sub>mac</sub> and HIV-2, respectively.

We have previously described the derivation of five genome-length proviral DNA clones of SIV<sub>sm</sub>/Delta/F236 (ref. 12). This virus was isolated in 1986 at the Delta Regional Primate Research Center from a rhesus macaque (F236) that died from opportunistic infections after experimental inoculation with SIV isolated from an asymptomatic sooty mangabey (E038). Mangabey E038 had been obtained from the Yerkes Regional Primate Research Center<sup>13</sup>. On transfection into H9 cells in culture, two of the five clones (λsmH-3 and λsmH-4) produced infectious virions. Progeny of both biologically active molecular clones established persistent infections in experimentally inoculated rhesus and pig-tail macaques (data not shown).

The complete nucleotide sequence (10,241 base pairs (bp)) of the infectious proviral molecular clone λsmH-4 was determined and analysed. The genomic organization of SIV<sub>sm</sub> was similar to SIV<sub>mac</sub> and HIV-2 (Fig. 1). The long terminal repeat (LTR) of SIV<sub>sm</sub> (815 bp) was 87% identical to SIV<sub>mac</sub> and 74%

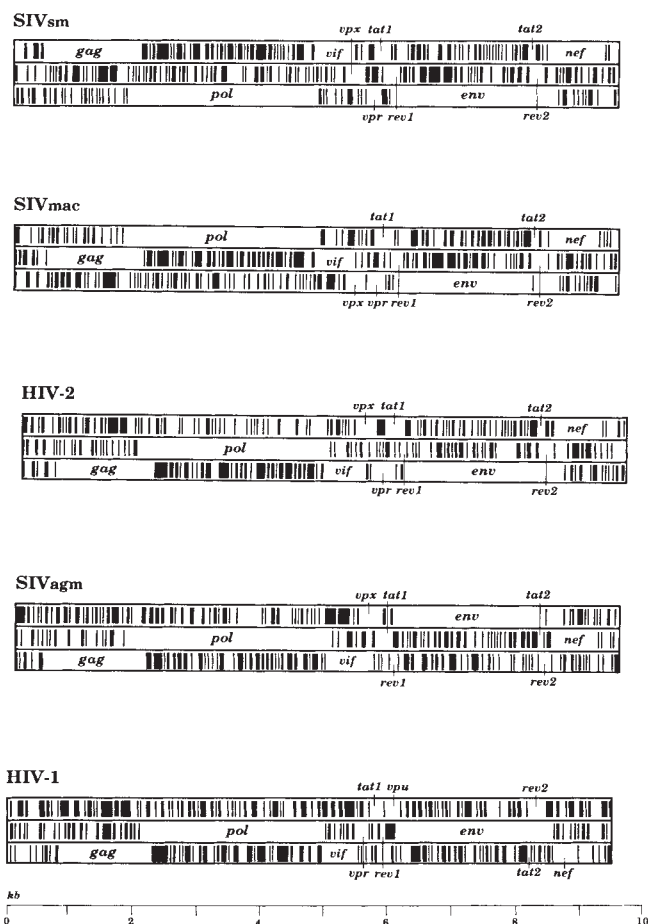


FIG. 1 Schematic representation of the open reading frames of five primate lentiviruses. Bars in each frame represent positions of stop codons. The sequences displayed are: SIV<sub>sm</sub> (strain F236); SIV<sub>mac</sub> (strain 142); HIV-2 (strain ROD); SIV<sub>agm</sub> (strain TYO-1); and HIV-1 (HXB2).

to HIV-2 overall. Regulatory sequences, including potential Sp1-binding regions and the TAR sequence, were highly conserved. Comparison of the deduced amino-acid sequences from the nine open reading frames in the SIV<sub>sm</sub> genome revealed significant similarities with SIV<sub>mac</sub> and HIV-2 (Table 1). The regions encoding Gag (Fig. 2) and Pol proteins were highly conserved, accounting for the serological cross-reactivity previously noted among the three viruses<sup>1</sup>. An interesting feature of the λsmH-4 *pol* nucleotide sequence was a single-base

TABLE 1 Amino-acid identity (%) between proteins of primate lentiviruses

|                                         | Gag | Pol | Vif | Vpr | Rev | Tat | Vpx | Envelope |       |       |                   |                  | Nef |
|-----------------------------------------|-----|-----|-----|-----|-----|-----|-----|----------|-------|-------|-------------------|------------------|-----|
|                                         |     |     |     |     |     |     |     | gp160    | gp120 | Whole | gp32/41           |                  |     |
|                                         |     |     |     |     |     |     |     |          |       |       | Before stop codon | After stop codon |     |
| SIV <sub>sm</sub> /Delta/F236 versus:   |     |     |     |     |     |     |     |          |       |       |                   |                  |     |
| SIV <sub>mac</sub> 142                  | 90  | 92  | 84  | 90  | 70  | 72  | 95  | 81       | 81    | 81    | 89                | 70               | 77  |
| HIV-2 (ROD)                             | 84  | 84  | 69  | 70  | 62  | 65  | 80  | 72       | 70    | 76    | 87                | 66               | 64  |
| SIV <sub>agm</sub> (TYO-1)              | 59  | 58  | 32  | —   | 43  | 32  | 38  | 46       | 45    | 47    | 56                | 36               | 35  |
| HIV-1 (HXB-2)                           | 55  | 57  | 28  | 45  | 32  | 26  | —   | 36       | 34    | 41    | 43                | 37               | 36  |
| SIV <sub>mac</sub> : 142 versus 251     | 96  | 97  | 90  | 98  | 88  | 88  | 97  | 92       | 91    | 94    | 94                | 93               | 82  |
| HIV-2: ROD versus NIH-Z                 | 92  | 91  | 86  | 85  | 86  | 85  | 84  | 80       | 78    | 84    | 89                | 78               | —   |
| SIV <sub>mac</sub> 142 versus HIV-2 ROD | 82  | 81  | 64  | 70  | 59  | 63  | 86  | 70       | 68    | 73    | 87                | 64               | 73  |

Sequences obtained from the Los Alamos HIV Sequence Database. Alignments performed using the PRTALN program<sup>23</sup> with default parameters. Amino-acid identities calculated with the inclusion of insertions and deletions. Dashed line indicates no comparison performed because: SIV<sub>agm</sub> lacks vpr; HIV-1 lacks vpx; or because of a large deletion in Nef of NIH-Z.



insertion near the putative proteolytic cleavage site for the integrase protein. This insertion introduced a frameshift and subsequent stop codon in the extreme 5' end of the integrase nucleotide sequence. The predicted amino-acid sequence of  $\lambda$ smH-3 in this region was analysed and found to be identical to SIV<sub>mac</sub> and HIV-2. It was interesting to note that after transfection of  $\lambda$ smH-4 into H9 cells in culture, reverse transcriptase activity was detected at 30 days, as opposed to 7 days for  $\lambda$ smH-3 (ref. 12). Experiments are in progress to define further the biological importance of this single-base insertion.

Genes in the complex central region were variably conserved (Table 1). Surprisingly, the essential regulatory genes, *rev* and *tat*, were relatively poorly conserved. However, distinctive structural motifs within *rev* and *tat* (for example, the arginine-rich domain in the second coding exon in *rev* and the cysteine cluster in the first coding exon of *tat*) were maintained among SIV<sub>sm</sub>, SIV<sub>mac</sub> and HIV-2. The most highly conserved central region gene is *vpx*, an open reading frame common to all characterized strains of SIV and HIV-2. The *vpx* gene product (p14<sup>vpx</sup>) is immunogenic in natural HIV-2 and SIV infection<sup>14,15</sup> and may bind to the single-stranded RNA genome in mature virions<sup>16</sup>.

Alignment of the predicted envelope-protein sequences of SIV<sub>sm</sub>, SIV<sub>mac</sub> and HIV-2 reveals an overall conservation of cysteine residues and other primary structural features (Fig. 3). Variable domains in *env*, predicted by analogy to HIV-1, were also variable among SIV<sub>sm</sub>, SIV<sub>mac</sub> and HIV-2. In addition, another highly variable region was identified within the putative cytoplasmic domain of the gp40 transmembrane subunit (residues 736–752). This area contains an in-frame stop codon that markedly truncates the cytoplasmic domain in some SIV<sub>mac</sub> and HIV-2 molecular clones<sup>2-4,17</sup>. The SIV<sub>sm</sub> *env* nucleotide sequence does not contain a premature stop-codon and is predicted to encode a full-length gp40, in contrast to the gp32 of SIV<sub>mac</sub> and several HIV-2 isolates<sup>18-20</sup>. This prediction was confirmed by the demonstration of a transmembrane protein of

relative molecular mass ( $M_r$ ) 40,000 on western blots of virions collected after transfection of the  $\lambda$ smH-4 and  $\lambda$ smH-3 clones into H9 cells<sup>12</sup>. The reduction in sequence similarity among the three envelope proteins after the stop codon region was significant (Table 1; Fig. 2). This may indicate either a lack of structural (or functional) constraints on the amino-acid sequence of the cytoplasmic domain or variable suppression or mutation of the stop codon resulting in a mixed population of viruses, or both. The full-length cytoplasmic domain seems to be synthesized during SIV<sub>mac</sub> infection of macaques, but is not detected in SIV<sub>mac</sub>-infected cells in tissue culture (ref. 21; M.M.-C., unpublished data). Thus, one possibility is that SIV<sub>mac</sub> and HIV-2 molecular clones containing the premature stop-

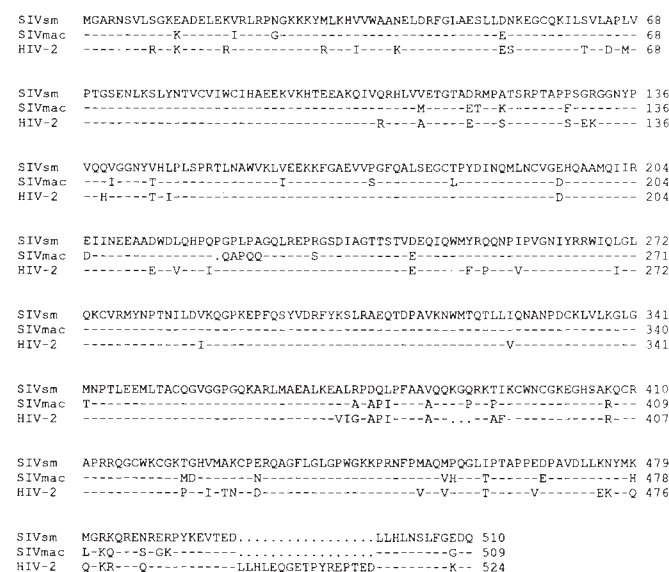


FIG. 2 Alignment of the predicted Gag protein sequences of SIV<sub>sm</sub>/Delta/F236, SIV<sub>mac</sub> (251) and HIV-2 (ROD). Dashes below the SIV<sub>sm</sub> sequence, identical amino acids; dots, gaps introduced to maximize the alignment.

**METHODS.**  $\lambda$ smH-4 was subcloned (in 1–3-kb fragments) into a plasmid vector (pTZ18R; Pharmacia) and sequenced by the dideoxynucleotide chain-terminating method, using T7 DNA polymerase (United States Biochemical). The *gag* nucleotide sequence was derived from subclone psmH-4B including the 5' non-coding *gag* sequence and a 5' portion of the *pol* gene. DNA sequences were analysed using the PC/Gen, IFIND, SEQ and GenALIGN programs (IntelliGenetics).

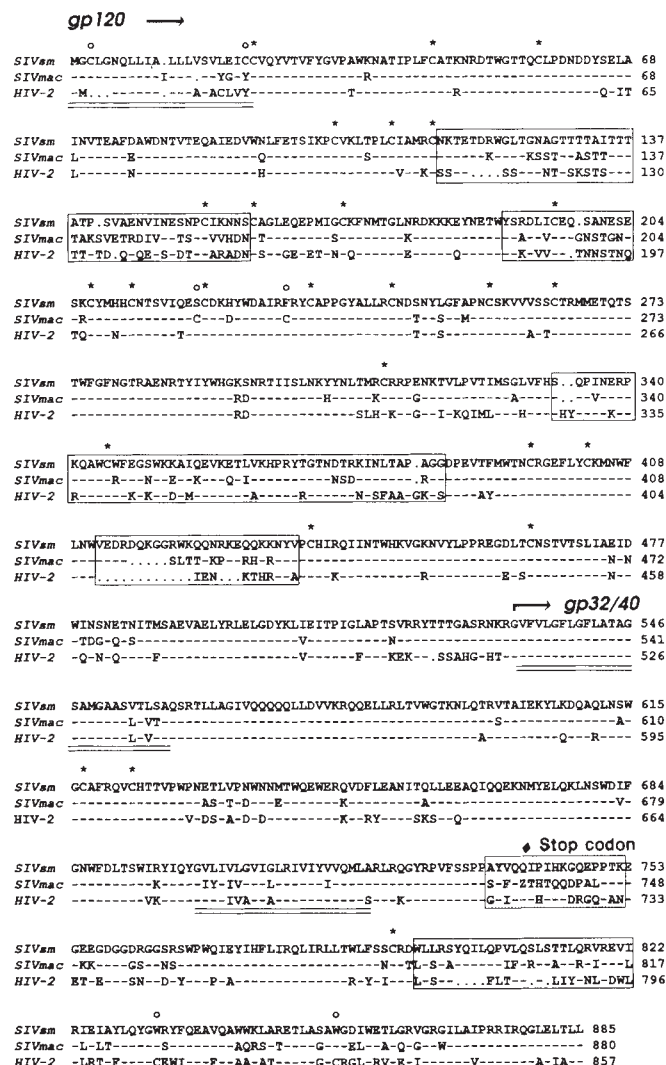
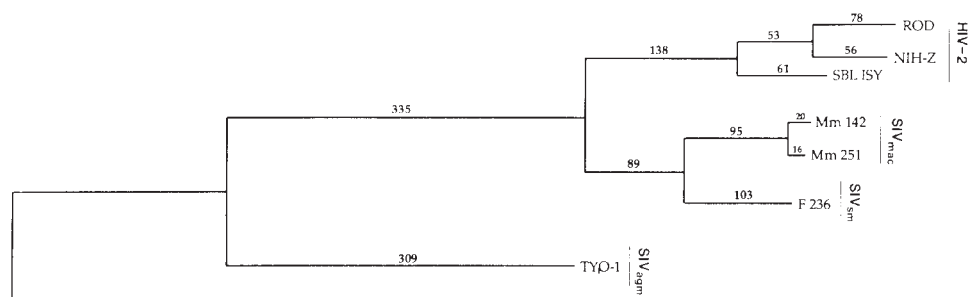


FIG. 3 Alignment of the predicted envelope glycoproteins of SIV<sub>sm</sub>/Delta/F236, SIV<sub>mac</sub> (251) and HIV-2 (ROD). Conventions for the alignment as in Fig. 1. Boxes, variable domains (as defined among HIV-1 isolates). In addition, another variable domain in gp40 surrounding the premature stop codon in SIV<sub>mac</sub> (black diamond) is enclosed by a shaded box. Cysteines conserved among all three sequences indicated by asterisks; non-conserved cysteines marked by open circles above the sequence. The putative proteolytic cleavage site between gp120 and gp40 is marked by an arrow. Predicted hydrophobic domains (signal, N terminus of gp40, and the membrane-spanning domain) are indicated by double underlining.

**METHODS.** The nucleotide sequence of the SIV<sub>sm</sub> *env* gene was determined (as described in Fig. 1) from three plasmid subclones of  $\lambda$ smH-4: psmH-4C2 (gp120); psmH-4D (gp40); and, psmH-4E (C terminus of gp40). DNA sequence analysis and alignments performed as in Fig. 1.

FIG. 4 Minimum-length evolutionary tree based on *gag* nucleotide sequences of the SIV/HIV-2 lentivirus subfamily.

**METHODS.** The tree was constructed from variation in the *gag* nucleotide sequence using PAUP (ref. 26) with the global branch-swapping option MULPARB. The total number of sites examined was 1,748, of which 858 were variable. The minimum length of the tree was 1,690 and the consistency index was 0.848. The tree was rooted on a random sequence (as in ref. 22). Lengths of the horizontal lines are proportional to the minimum number of single nucleotide substitutions required to generate the observed variation (also indicated by the numeric



branch lengths above each line). The length of the vertical lines is for clarity only. Evolutionary tree analyses based on deduced protein sequences<sup>27</sup> generated a similar branching order of divergence.

codon represent viruses artificially selected for their ability to grow in tissue culture. We are investigating the functional importance of this highly variable region.

A distinctive feature of the *SIV<sub>sm</sub>* sequence is an extended open reading frame for *nef*, the negative regulatory gene. In HIV-1, HIV-2 and *SIV<sub>mac</sub>* molecular clones, the *nef* gene is variable in length; in general, the HIV-1 *nef* gene is truncated relative to *SIV* and HIV-2. The *SIV<sub>sm</sub>* *nef* gene contains an apparent frameshift (single-base insertion) near the C terminus that extends the predicted reading frame for an additional 54 amino acids relative to *SIV<sub>mac</sub>* and HIV-2. The sequences of both *SIV<sub>sm</sub>* biologically active clones are identical through this region, indicating that the extended reading frame is not limited to a single clone.

To investigate further the relationship of *SIV<sub>sm</sub>* to *SIV<sub>mac</sub>* and HIV-2, we constructed minimum-length evolutionary trees<sup>22</sup> based on the nucleotide variations in the LTR and the *gag* and *env* genes (the *gag* nucleotide tree is shown in Fig. 4). The trees are all consistent in the branching order of divergence, although branch lengths vary (*env* genes are more divergent). By this analysis of *gag*, *SIV<sub>sm</sub>* was at least 12% different from *SIV<sub>mac</sub>* and 26–28% different from HIV-2. A precise time calibration for this tree was not possible because the mutation rates for *SIV<sub>mac</sub>* and HIV-2 have not been determined and could not be assumed to be equal or constant. However, if we assume a representative mutation rate for the *gag* gene of 0.5% per virus per year (estimated from HIV-1 (ref. 22), *SIV<sub>sm</sub>* and *SIV<sub>mac</sub>* might have diverged from each other 10 years ago (minimum), and *SIV<sub>sm</sub>*/*SIV<sub>mac</sub>* might have diverged from HIV-2 as recently as 30 years ago. The latter figure is in agreement with previously published data<sup>23</sup>. We emphasize that these calculations are only rough estimates and are based on mutation rates that have not been experimentally determined for SIV or HIV-2.

These evolutionary estimates are consistent with the limited information regarding the origins of *SIV<sub>sm</sub>*, *SIV<sub>mac</sub>* and HIV-2. First, consider the relationship of *SIV<sub>sm</sub>* and *SIV<sub>mac</sub>*. Because Asian macaques are not infected with SIV in the wild, captive macaques probably acquired SIV by cross-species transmission from an SIV-infected African primate. SIV infection of captive macaques has been traced back to the early 1970s at the New England Regional Primate Research Center; the putative index cases were macaques obtained from another primate facility (N. Letvin, personal communication). Although not documented, these macaques may have been housed in captivity with African primates (sooty mangabeys or others) before arrival in New England. If *SIV<sub>mac</sub>* 251 (the published prototype *SIV<sub>mac</sub>*) is thought of as a 1982 virus (the year that rhesus macaque 251 died)<sup>24</sup>, then we estimate that *SIV<sub>sm</sub>* and *SIV<sub>mac</sub>* might have diverged in the late 1960s or early 1970s (based on the minimum divergence time of 10 years, estimated above). In support of this hypothesis, *SIV<sub>sm</sub>* has probably been in the United States since before 1968, the last year that wild-caught West African

mangabeys were added to the Yerkes colony<sup>11</sup>. Therefore, the molecular and epidemiological data together suggest that an SIV-infected sooty mangabey (or closely related species) was the source of *SIV<sub>mac</sub>*.

The unusually close relationship of SIV from captive macaques with the human lentivirus HIV-2 has remained an enigma. We propose that our data on the genetic relatedness of *SIV<sub>sm</sub>*, *SIV<sub>mac</sub>* and HIV-2 provide the previously missing link. The relationship between HIV-2 and *SIV<sub>sm</sub>* can be summarized by three observations: first, SIV-infected sooty mangabeys inhabit West Africa; second, HIV-2 infection of humans is endemic in West Africa; and, third, the genomes of *SIV<sub>sm</sub>* and HIV-2 are closely related. Clearly, the *SIV<sub>sm</sub>* that we have cloned and sequenced is not the direct ancestor of HIV-2 or *SIV<sub>mac</sub>*. Our sequence represents an *SIV<sub>sm</sub>* that has probably been in North America for more than 20 years. Also, the HIV-2 sequences reported to date probably represent 20 or more years of evolution of the virus in humans in West Africa. Therefore, the degree of sequence similarity we report here is remarkable and argues that these viruses evolved fairly recently from a common ancestor. A plausible interpretation of these data is that in the past 30–40 years SIV from a West African sooty mangabey (or closely related species) successfully infected a human and evolved as HIV-2. Our data cannot exclude the possibility that HIV-2 from a human was passed to a sooty mangabey and subsequently evolved as *SIV<sub>sm</sub>*. Sequences of older HIV-2 and SIV isolates (from mangabeys and other species) are required to resolve these issues. Characterization of novel SIVs from primates in different regions of Africa may provide further clues to the origins of human lentiviruses. □

Received 27 February; accepted 18 April 1989.

- Schneider, J. & Hunsmann, G. *AIDS* **2**, 1–9 (1988).
- Chakrabarti, L. et al. *Nature* **328**, 543–547 (1987).
- Franchini, G. et al. *Nature* **328**, 538–543 (1987).
- Hirsch, V., Reidel, N. & Mullins, J. I. *Cell* **49**, 307–319 (1987).
- Fukasawa, M. et al. *Nature* **333**, 457–461 (1988).
- Tsujimoto, H. et al. *J. Virol.* **62**, 4044–4055 (1988).
- Daniel, M. D. et al. *Science* **228**, 1201–1204 (1985).
- Ohta, Y. et al. *Int. J. Cancer* **41**, 115–122 (1988).
- Lowenstine, L. J. et al. *Int. J. Cancer* **38**, 563–574 (1986).
- Murphy-Corb, M. et al. *Nature* **321**, 435–437 (1986).
- Fultz, P. N. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5286–5290 (1986).
- Hirsch, V. M. et al. *J. med. Primatol.* (in the press).
- Baskin, G. B., Murphy-Corb, M., Watson, E. A. & Martin, L. N. *Vet. Path.* **25**, 456–467 (1988).
- Yu, X.-F., Ito, S., Essex, M. & Lee, T.-H. *Nature* **335**, 262–265 (1988).
- Franchini, G., Ruschke, J., O'Keefe, T. & Wong-Staal, F. *AIDS Res. Hum. Retroviruses* **4**, 243–259 (1988).
- Henderson, L. E., Sowder, R. C., Copeland, T. D., Benveniste, R. E. & Oroszlan, S. *Science* **241**, 199–201 (1988).
- Guyader, M. et al. *Nature* **326**, 662–669 (1985).
- Barin, F. et al. *Lancet* **ii**, 1387–1389 (1985).
- Albert, J. et al. *AIDS Res. Hum. Retroviruses* **3**, 3–10 (1987).
- Clavel, F. et al. *Science* **233**, 343–346 (1986).
- Franchini, G., Kanaki, P. J., Bosch, M., Fargnoli, K. & Wong-Staal, F. *AIDS Res. Hum. Retroviruses* **4**, 251–258 (1988).
- Smith, T. F., Srinivasan, A., Schochetman, G., Marcus, M. & Myers, G. *Nature* **333**, 573–575 (1988).
- Sharp, P. M. & Li, W.-H. *Nature* **336**, 351 (1988).
- Hunt, R. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5085–5089 (1983).

25. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726–730 (1983).  
 26. Day, W. H. E. *J. theor. Biol.* **103**, 429–438 (1983).  
 27. Feng, E. F. & Doolittle, R. *J. molec. Evol.* **25**, 351–360 (1987).

ACKNOWLEDGEMENTS. The authors thank Drs G. Myers and R. Doolittle for evolutionary tree analyses and discussions, G. Dapolito, C. McGann and S. Kitov for technical assistance, and Dr R. Chanock for discussions and support. The nucleotide sequence described in this paper has been deposited in the GenBank/EMBL database under the accession number X14307. The entire sequence can be obtained through GenBank/EMBL or by writing to the authors.

## Binding of labelled influenza matrix peptide to HLA DR in living B lymphoid cells

Ruggero Ceppellini\*, Guido Frumento\*,  
 Giovanni B. Ferrara\*†, Roberto Tosi‡,  
 Alberto Chersi§ & Benvenuto Pernis||

\* Istituto Nazionale per la Ricerca sul Cancro, Genova, Italy

† Cattedra di Genetica Umana, Università di Napoli and Centro Ricerche Immunopatologiche, Avis, Bergamo, Italy

‡ Istituto di Biologia Cellulare, CNR, Roma, Italy

§ Ripartizione di Oncologia Sperimentale, Istituto Regina Elena, Roma, Italy

|| Departments of Microbiology and Medicine, Columbia University, New York, New York 10032, USA

**T CELLS recognize protein antigens as fragments (peptides) held in a defined binding site<sup>1</sup> of class I or class II major histocompatibility (MHC) molecules. The formation of complexes between various immunologically active peptides and different MHC molecules has been demonstrated directly in binding studies between the peptides and solubilized, purified molecules of class II MHC<sup>2–6</sup>. Studies with intact cells, living or fixed, have not directly demonstrated the binding of the peptides to MHC molecules on antigen-presenting cells, but the formation of such complexes has been shown indirectly through the capacity of antigen-presenting cells to stimulate specific T cells<sup>7–9</sup>. Here we report evidence that supports directly the binding of radiolabelled influenza matrix peptide 17–29 to products of the human class II MHC locus HLA-DR, on living homozygous B-cell lines, and we show that the kinetics of such binding is much faster with living cells than with fixed cells. Furthermore, whereas the peptide reacts with HLA-DR molecules of all alleles, it binds preferentially to DR1, the restricting element in antigen presentation.**

The results of direct binding experiments with <sup>125</sup>I-labelled peptide 17–29 of influenza virus matrix protein, are reported in Fig. 1. These concern 21 HLA-homozygous Epstein-Barr virus-transformed B-cell lines from the tenth Histocompatibility Workshop, plus two human resting polyclonal T-cell populations that do not express class II MHC, and two human T-cell lines, one of which (MOLT 4) does not express class II MHC, and the other (HUT 78) expresses class II (DR4, DRw53, DQw3) at a level comparable to that of the B-cell lines, as shown by fluorescence-activated cell sorter analysis with monoclonal antibody Q5/13. The MHC types of the HLA homozygous B-cell lines are given in Table 1.

The data shown in Fig. 1 indicate that peptide 17–29 of influenza matrix binds within 30 min and in reproducible amounts to living human B cells. Cells that carry the DR1 allele, the restricting element for presentation of peptide 17–29 to T cells<sup>10,11</sup>, show increased binding. All of our DR1-expressing cell lines also carried DQ5, which is in linkage disequilibrium with DR1, but we had one DQ5-expressing line (SP00 10) that did not express DR1, and this was not in the high-binding group. The high-binding group did however include one line that does

not carry DR1—the WDV line expressing DR w13 (note that not all lines expressing Dr w13 have this property, see line WT47). All B-cell lines bound much more peptide than T cells, with an intermediate value given by HUT 78, a T-cell line that expresses class II MHC. The data shown in Fig. 1 are compatible with the possibility that peptide 17–29 of influenza matrix binds to all DR molecules, with preference for its restricting allele product, DR1.

More compelling evidence for this possibility is given by the data summarized in Fig. 2. These show that the binding of the peptide, both to a high-binder and to a low-binder B-cell line, is inhibited by an antibody for class II MHC and is not affected by an anti-class I. We obtained similar results (data not shown) with two different monoclonal anti-class I and anti-class

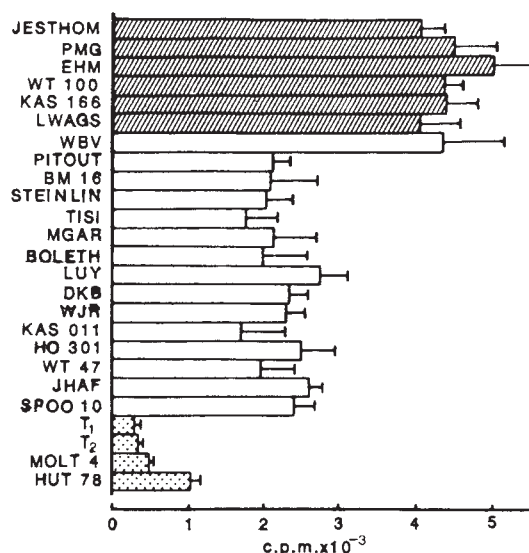


FIG. 1 Binding of influenza matrix peptide 17–29-Tyr <sup>125</sup>I to living human B and T cells. The data are the mean c.p.m.  $\pm$  1 s.d., from 18 separate experiments for EHM cells, 9 experiments for JHAF cells, 6 for BM 16 and MOLT 4, and triplicate for the other cell lines or populations. ■, DR1-positive B-cell lines; □, DR1-negative B-cell lines; ▨, T-cell lines (MOLT 4 and HUT 78) or polyclonal T-cell populations. Resting T cells (T<sub>1</sub> and T<sub>2</sub>) were derived from the blood of two healthy donors by the E-rosette procedure.

**METHODS.** Peptide 17–29 of the influenza virus matrix protein (Ser-Gly-Pro-Leu-Lys-Ala-Glu-Ile-Ala-Gln-Arg-Leu-Glu) was synthesized by the solid-phase method<sup>19</sup> using a Du Pont-Vega Coupler, Model 100. An additional tyrosine was added at the C-terminal end of the peptide to allow iodination. *N*-t-butoxycarbonyl protection was used, and trifunctional amino acids were protected as follows: Ser (*O*-benzyl); Lys (2-chloro-benzyl-oxy-carbonyl); Glu (*O*-benzyl); Arg (tosyl); Tyr (2,6-dichloro-benzyl). After deprotection and cleavage from the support by treatment with trifluoromethanesulphonic acid (40 min at 40 °C) (ref. 20), the free peptide was purified by gel filtration on a 2 × 100 cm column of Sephadex G25 (superfine) in 0.01 M ammonia. Thin-layer chromatography showed that the peptide was substantially homogenous, and amino-acid analysis indicated that the composition was within 5 per cent of the theoretical value. The peptide was iodinated by the chloramine-T method<sup>21</sup> to a specific activity of 50  $\mu$ Ci  $\mu$ g<sup>-1</sup> in the presence of an equimolar concentration of cold KI. Incorporation was 70–80%. Free <sup>125</sup>I was removed by gel filtration on Sephadex G-10. The labelled peptide was kept at 4 °C in 0.075 M Tris, pH 7.8, containing 2% ethanol and 0.001% sodium azide. To achieve the working concentration the peptide was diluted in complete medium immediately before testing. Experiments were performed within 3 weeks from the labelling. Binding test: Cells were washed once in RPMI 1640 and resuspended in the same medium supplemented with nutridoma HU (Boehringer), 2 mM L-glutamine and 1 mM pyruvate at a concentration of  $2 \times 10^7$  cells per ml. Cell suspension (50  $\mu$ l) was added to an equal volume of peptide solution, containing 10 ng labelled peptide in 15 ml centrifuge tubes. If not otherwise indicated, after incubation at 37 °C in CO<sub>2</sub> incubator for 30 min, the cells were washed three times at 4 °C and the pellet was counted in a Packard 5110 gamma counter to assay the cell-associated radioactivity.

|| To whom correspondence should be addressed.



II antibodies (PTE 29.12 and PTF 22.1). Furthermore, a comparable inhibition with either cell line is obtained with an antiserum (FE 204 or LH 986) that reacts with the DR allele product of the line itself, thus favouring the view that the majority of influenza matrix peptide 17-29 that binds to human B cells binds to DR molecules and not to the products of other class II sub-regions.

Figure 3 gives the binding of matrix peptide 17-29 to B- and T-cell lines as a function of peptide concentration. The dynamic state of class II MHC molecules on the membrane of living B cells<sup>12,13</sup> means that it is not possible to calculate a precise  $K_D$  value. But, using peptide 17-29 and purified DR1 molecules, a  $K_D$  value of  $10^{-7}$  M has been estimated (T. Jardetsky, personal communication). The data summarized in Fig. 3 would indicate

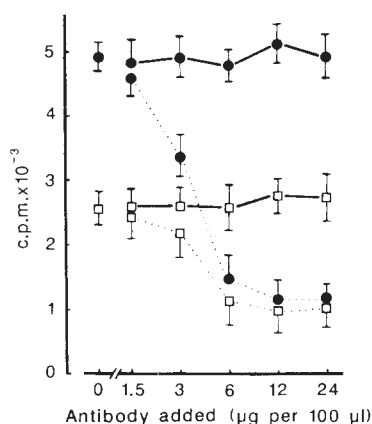


FIG. 2 Inhibition of binding of influenza matrix-peptide 17-29 Tyr<sup>125</sup>I on two B-cell lines (line EHM (●) and line MGAR □) by anti-class II MHC antibodies. Conditions as in Fig. 1, but incubation was also performed in the presence of purified monoclonal anti-DR antibody (dotted line) or anti-class I antibody (continuous line). The anti-DR antibody S 1.5 (ref. 22) (donated by M. Trucco) was purified from the culture supernatant with Affigel Protein A MAPS 11 (Bio Rad). The anti-class I MHC antibody was W 6/32, obtained from the American Type Culture Collection (Rockville) and purified from the culture supernatant with Protein A Sepharose CL4B (Pharmacia). Both antibodies were dialysed before use and added to the cells in the stated concentration 30 min before the labelled peptide. The results are the average of triplicate cultures  $\pm$  1 s.d.

TABLE 1 HLA specificities of the B-cell lines

| Name       | A     | B   | C   | DR  | DP    | DQ |
|------------|-------|-----|-----|-----|-------|----|
| JESTHOM    | 2     | 27  | w1  | 1   | w1    | w5 |
| PMG075     | 3, 33 | w65 | w8  | 1   | w3, 4 | w5 |
| EHM        | 3     | 35  | w4  | 1   | w4    | w5 |
| WT 100 BIS | 11    | 35  | w4  | 1   | w1    | w5 |
| KAS 116    | 24    | 51  | —   | 1   | —     | w5 |
| LWAGS      | w33   | 14  | w8  | 1   | w3, 4 | w5 |
| WDV        | 2     | 38  | —   | w13 | w4    | w6 |
| PITOUT     | 29    | 44  | —   | 7   | w4    | w2 |
| BM16       | 2     | 18  | w7  | w12 | w2    | w7 |
| STEINLIN   | 1     | 8   | —   | 3   | w3, 4 | w2 |
| TISI       | 24    | 35  | w4  | w11 | —     | w7 |
| MGAR       | 26    | 8   | w7  | w15 | w4    | w6 |
| BOLETH     | 2     | w62 | w10 | 4   | w4    | w8 |
| LUY        | 2     | 51  | —   | w8  | w1, 4 | w7 |
| DKB        | 24    | w60 | w10 | 9   | w4    | w9 |
| WJR 076    | 2     | w57 | w7  | w16 | w3, 4 | w6 |
| KAS 011    | 1     | 37  | w6  | 2   | w2    | w1 |
| HO 301     | 3     | 14  | w8  | w6  | w5    | w6 |
| WT47       | 32    | 44  | w5  | w13 | w2    | w6 |
| JHAF       | 31    | 51  | —   | 4   | w3    | w7 |
| SP00 10    | 2     | 44  | w5  | w11 | w2    | w5 |

an affinity in that range. It is therefore possible that the intrinsic affinity of a given class II molecule for a given peptide is the same whether the MHC molecule is isolated or on the membrane of a cell.

Certainly, the affinity of matrix 17-29 for DR1 is higher than in other systems<sup>2,3</sup>, which may explain how it has been possible to detect the binding of a peptide to class II MHC on the membrane of living cells above a background of binding which is not due to MHC molecules. This non-specific binding represents 25% of total binding as assessed by the aliquot that cannot be inhibited by an excess of anti-class II antibody (Fig. 2). Non-specific binding was also measured by determining the aliquot that could not be inhibited with a 100-fold excess of cold ligand<sup>14</sup>, which by this method was found to represent 29% of the total.

Some peptide-binding proteins, in addition to MHC and receptors for peptide hormones, have been described in different cells<sup>15,16</sup>. Figure 3a shows binding of matrix peptide 17-29 to MOLT 4 cells that do not express class II MHC; the binding to MOLT 4 is of lower affinity than the binding to EHM cells (for MHC type, see Table 1). If similar non-class II-dependent binding also takes place with B cells, it might be equal to or higher than the binding of the peptide to class II MHC in systems of lower affinity than that of peptide 17-29 for DR1.

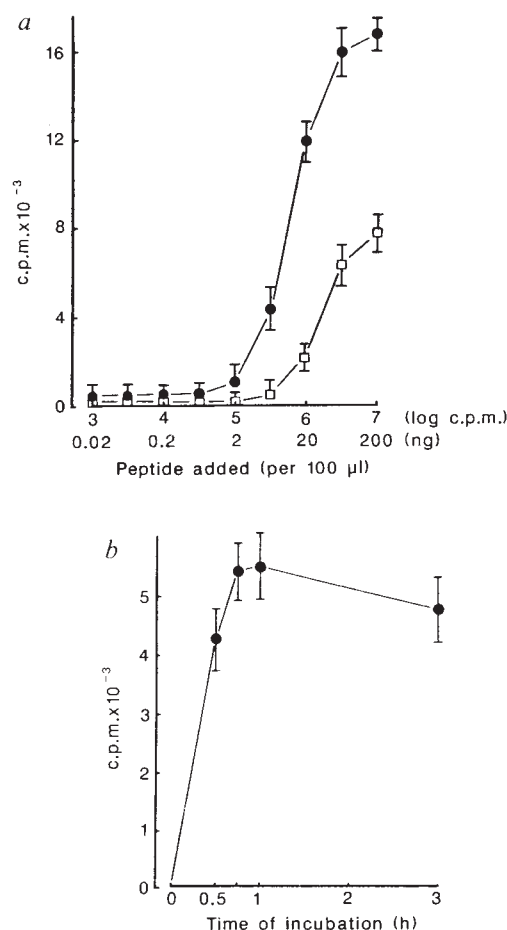


FIG. 3 a, Log<sub>10</sub> saturation isotherm of influenza matrix peptide 17-29 Tyr<sup>125</sup>I with EHM (DR1) cells (●) and with MOLT 4 (Class II negative) cells (□). Data obtained by incubation of  $10^6$  cells for 30 min at 37 °C with indicated amounts of peptide in 100  $\mu$ l medium followed by three washings. Each point gives the mean of triplicate cultures  $\pm$  1 s.d. b, Kinetics of binding of influenza matrix peptide 17-29-Tyr<sup>125</sup>I to EHM cells.  $10^6$  cells were incubated with 10 ng peptide for indicated period of time in 100  $\mu$ l medium at 37 °C, then washed three times. Each point gives the mean of triplicate cultures  $\pm$  1 s.d.

This might explain the problems so far encountered in the studies of binding labelled peptides to MHC on intact cells<sup>2,8</sup>. At the concentration of peptide 17-29 (100 ng ml<sup>-1</sup>) that we have routinely used, the ratio of binding to DR1 versus the other binding is favourable. The data summarized in Fig. 3a also indicate that the binding of peptide 17-29 to EHM cells approaches saturation at a level of class II-dependent binding of about  $7 \times 10^4$  molecules per cell; at this level ~1% of the peptide offered is bound. This binding involves at most 15% of all DR molecules that are found on the membrane of an EBV-transformed B-cell line<sup>17</sup>, which is in agreement with the fact that the majority of class II MHC binding sites are occupied by endogenous peptides<sup>18</sup>.

Figure 3b shows the kinetics of binding of peptide 17-29 to EHM cells; maximum binding is achieved in 45 min, which is much shorter than the 6 hours or longer needed for maximum binding of peptides to purified class II MHC molecules<sup>2,3</sup>. In fact, in the same system of peptide 17-29 and DR1, the binding to isolated DR1 molecules is not even complete in 6 hours (T. Jardetsky, personal communication). It would therefore seem that the rapid binding to DR1 of influenza matrix peptide 17-29 seen with living cells is related to their biological activity and is not simply the consequence of the intrinsic high affinity of the system. Roosneck *et al.*<sup>9</sup> have seen that living B cells become fully competent to present a tetanus toxoid peptide to T cells after 40 minutes of exposure to it. Support for the concept that some activity of the living cells is necessary for the rapid loading of class II MHC with a peptide comes from experiments with glutaraldehyde-fixed EHM cells (Fig. 4), showing that these cells have lost the capacity for rapid binding of matrix peptide 17-29.

Fixation does not eliminate the class II MHC-dependent binding to the same cells when these are exposed to the peptide for 18 hours. This is in agreement with the previous report of Shimonkevitz *et al.*<sup>7</sup> showing that glutaraldehyde-fixed cells can stimulate T cells in a specific way when incubated in the presence of a peptide for 24 hours. It seems that the peptide-binding properties of class II MHC molecules in fixed cells are similar to those of the same molecules after purification. Figure 4 also shows that peptide binding by intact cells does not occur in the cold.

In conclusion, it seems that living B cells have a system for relatively rapid loading of class II MHC molecules with peptides. This system is probably not only operative at the level of the cell membrane, but also in intracellular vesicles where peptides derived from processing of internalized protein antigens are present. In this case, the MHC loading time must be shorter than the intracellular transit time of either the newly synthesized or the recycling class II MHC molecules. The molecular

associations and structural events that make the rapid MHC/peptide associations in living B cells possible remain to be investigated. □

Received 6 December 1988; accepted 10 April 1989.

1. Bjorkman, P. *et al.* *Nature* **329**, 512-519 (1987).
2. Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. R. *Nature* **317**, 359-361 (1985).
3. Buus, S., Sette, A., Colon, S. M., Jervis, D. M. & Grey, H. M. *Cell* **47**, 1071-1077 (1986).
4. Guillet, J. G., Lai, M. Z., Briner, T. J., Smith, J. A. & Gefter, M. L. *Nature* **324**, 260-262 (1987).
5. Watts, T. H. & McConnell, H. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9660-9664 (1986).
6. Luescher, I. F., Allen, P. M. & Unanue, E. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 871-874 (1988).
7. Shimonkevitz, R., Kappler, J., Marrack, P. & Grey, H. *J. exp. Med.* **158**, 303-316 (1983).
8. Shimonkevitz, R., Colon, S., Kappler, J. W., Marrack, P. & Grey, H. M. *J. Immun.* **133**, 2067-2074 (1984).
9. Roosneck, E., Demotz, S., Corradin, G. & Lanzavecchia, A. *J. Immun.* **140**, 4079-4082 (1988).
10. Eckels, D. *et al.* *Immunogenetics* **19**, 409-423 (1984).
11. Rothbard, J. B. *et al.* *Cell* **52**, 515-523 (1988).
12. Pernis, B. *Immunol. Today* **6**, 45-49 (1985).
13. Harding, C. V. & Unanue, E. R. *J. Immun.* **142**, 12-19 (1989).
14. Limbird, L. *Cell Surface Receptors* (Nijhoff, Boston, 1986).
15. Nairn, R., Spengler, M. L., Hoffman, M. D., Solvay, M. J. & Thomas, D. W. *J. Immun.* **133**, 3225-3234 (1984).
16. Lakey, E. K., Margoliash, E. & Pierce, S. K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1659-1663 (1987).
17. Trucco, M., de Petris, S., Garotta, G. & Ceppellini, R. *Human Immunol.* **3**, 233-243 (1980).
18. Buus, S., Sette, A., Colon, S. M. & Grey, H. M. *Science* **242**, 1045-1047 (1988).
19. Merrifield, R. B. *J. Am. chem. Soc.* **85**, 2149-2154 (1963).
20. Yajima, H. *et al.* *Chem. Pharm. Bull.* **23**, 371-374 (1975).
21. Greenwood, F. C., Hunter, W. M. & Glover, J. S. *Biochem. J.* **89**, 114-123 (1963).
22. Trucco, M. M., Garotta, G., Stocker, J. W. & Ceppellini, R. *Immunol. Rev.* **47**, 237-242 (1979).

ACKNOWLEDGEMENTS. We thank Ted Jardetsky for communicating the results of unpublished work, Wayne Hendrickson for reading the manuscript, S. Ferrone for antibody Q5/13, G. Damiani for monoclonal antibodies PTE 29.12 and PTF 22.1, and H. Festenstein for anti-DR4 antiserum LH 986 Hugues. This work was supported by the National Research Council of Italy (C.N.R.) and by the NIH.

## A recombinant immunotoxin consisting of two antibody variable domains fused to *Pseudomonas* exotoxin

Vijay K. Chaudhary\*, Cary Queen†, Richard P. Junghans‡, Thomas A. Waldmann‡, David J. FitzGerald\* & Ira Pastan\*

\* Laboratory of Molecular Biology, DCBD, and † Metabolism Branch, DCBD, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

‡ Protein Design Labs, 3181 Porter Drive, Palo Alto, California 94304, USA

ANTIBODIES and growth factors have been chemically coupled to different toxins to produce cytotoxic molecules that selectively kill cells bearing appropriate antigens or receptors<sup>1,2</sup>. Antibody-toxin conjugates (immunotoxins) produced using conventional chemical coupling techniques have several undesirable characteristics. The smallest binding unit of an antibody is an Fv fragment which consists of a light and heavy chain variable domain. Recently, active single chain Fv fragments of antibodies have been produced in *Escherichia coli* by attaching the light and heavy chain variable domains together with a peptide linker<sup>3,4</sup>. Here we describe the construction and expression in *E. coli* of a single chain antibody toxin fusion protein, anti-Tac(Fv)-PE40, in which the variable regions of anti-Tac, a monoclonal antibody to the p55 subunit of the human interleukin-2 receptor<sup>5</sup>, are joined in peptide linkage to PE40, a modified form of *Pseudomonas* exotoxin lacking its binding domain. Anti-Tac(Fv)-PE40 was very cytotoxic to two interleukin-2 receptor-bearing human cell lines but was not cytotoxic to receptor-negative cells.

Immunotoxins made by chemically attaching a toxin to an intact antibody contain the constant region of the antibody, which is not necessary for immunotoxin action, but which reduces its access to target cells outside the circulation, and increases its immunogenicity. Furthermore, the product is

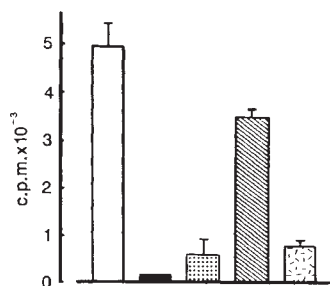


FIG. 4 Effect of glutaraldehyde fixation on binding of influenza matrix peptide 17-29-Tyr<sup>125</sup>I. Cells were fixed with glutaraldehyde (Serva, Heidelberg) exactly as indicated by Shimonkevitz *et al.*<sup>7</sup>. They were then incubated for indicated periods of time at 37 °C with 10 ng peptide (in 0.1 ml medium with  $10^6$  cells). ▨, EHM cells, incubation 30 min; ▨, EHM cells incubation 18 h; ▨, MOLT 4 cells, incubation 18 h. In addition, the figure gives the data for unfixed EHM cells incubated for 30 min at 37 °C (open bar) or at 4 °C (filled bar). Every column gives the mean of triplicate cultures  $\pm 1$  s.d.

heterogeneous and the yields are often poor. Recombinant DNA techniques have been used to produce chimaeric growth factor-toxin fusion proteins in *E. coli*<sup>6-10</sup>. These molecules can be readily purified in large amounts and contain only the sequences required for specific cell recognition and the cytotoxic activity of the toxin. Because the antigen-binding site of an antibody is composed of two separate polypeptide chains, it has been difficult to produce antibody-toxin chimaeric proteins in *E. coli*. To create a single-chain recombinant immunotoxin, we assembled a plasmid, pVC70108, which contains a 348-base pair (bp) DNA segment encoding an anti-Tac heavy chain variable domain (VH) joined to a 318-bp DNA segment encoding a light chain variable domain (VL) by a 45-bp linker; VL was in turn joined to a DNA segment encoding amino acids 253-613 of *Pseudomonas* exotoxin (PE) (Fig. 1). (The cloning and sequence of the anti-Tac variable regions will be described elsewhere.) The assembled gene is under the control of a T7 promoter<sup>11</sup>. The authenticity of the coding region of the plasmid was confirmed by DNA sequencing (data not shown). Upon induction with isopropyl- $\beta$ -D-thiogalactoside (IPTG), BL21 ( $\lambda$ DE3) carrying plasmid pVC70108 produced large amounts of a protein of relative molecular mass ( $M_r$ ) ~65,000 (65 K), as shown by SDS-PAGE (lane 1, Fig. 2b). On immunoblots, the 65 K chimaeric protein reacted with an antibody to PE (data not shown). The fusion protein was mostly contained in the 100,000 g pellet (lane 3, Fig. 2b) of the sonicated spheroplasts (lane 2, Fig. 2b). This pellet was used as the source to prepare anti-Tac(Fv)-PE40. Guanidine hydrochloride denaturation followed by rapid dilution was used to solubilize and renature the chimaeric protein<sup>8</sup>. The renatured protein was applied to a Mono Q column and the monomeric form of the fusion protein eluted at 0.2-0.22 M NaCl (Fig. 2b, lane 4 and Fig. 2a). High-molecular weight aggregates were eluted at higher ionic strength (fractions 42-50; Fig. 2a). Further purification of the chimaeric protein was carried out on a TSK-250 gel filtration column; the chimaeric protein eluted as a symmetrical peak at the location expected

for a 65 K protein (data not shown). SDS-PAGE showed the protein to be >95% pure (lane 5, Fig. 2b) and N-terminal amino acid analysis showed the protein had the expected sequence, Met-Gln-Val-. Highly purified monomeric anti-Tac(Fv)-PE40 (~200  $\mu$ g) was obtained from 1 litre of cells grown to an optical density at 650 nm of 0.6 before induction.

The anti-Tac antibody binds to the p55 subunit (Tac antigen, low affinity receptor) of the interleukin-2 (IL-2) receptor, which is present in large amounts on HUT-102 cells<sup>5</sup>. Therefore, the chimaeric protein was initially tested for cytotoxicity on HUT-102 cells. Anti-Tac(Fv)-PE40 inhibited protein synthesis in a dose-dependent manner with a 50% inhibitory dose ( $ID_{50}$ ) of 0.15 ng ml<sup>-1</sup> ( $2.3 \times 10^{-12}$  M) in a 20 h assay (Fig. 3; Table 1). At concentrations >4 ng ml<sup>-1</sup>, there was complete inhibition of protein synthesis. Several specificity controls were carried out. Addition of excess anti-Tac (10  $\mu$ g ml<sup>-1</sup>) prevented the cytotoxicity of anti-Tac(Fv)-PE40 on HUT-102 cells, whereas a control monoclonal antibody, OVB3, directed against an antigen found on ovarian cancer cells<sup>12</sup> did not (Fig. 3). The human T-cell leukaemia line Cr II.2 (ref. 13), which has a lower number

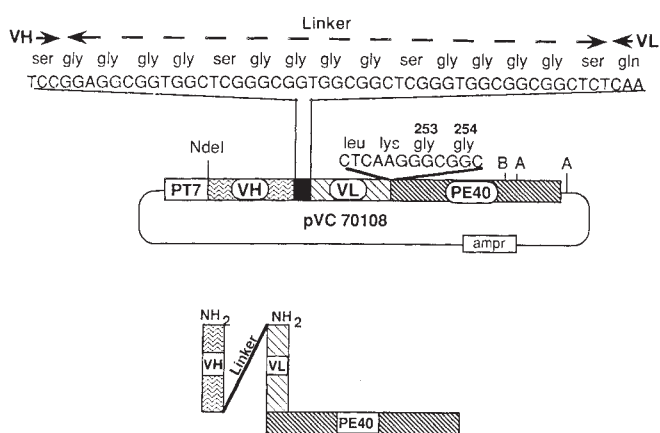


FIG. 1 a, Expression plasmid for anti-Tac(Fv)-PE40. Expression plasmid pVC70108 contains a fusion gene encoding various domains of anti-Tac (the variable domain of the heavy chain (VH, first 116 amino acids of mature heavy chain), a 15-amino-acid linker ((Gly-Gly-Gly-Gly-Ser)<sub>3</sub>), and the variable domain of the light chain (VL, first 106 amino acids of mature light chain)) and amino acids 253-613 of PE (refs 6-9, 17) as a single polypeptide chain. The gene is under control of a T7 promoter linked to a Shine-Dalgarno sequence and initiation codon (PT7) as described previously<sup>6-9,17</sup>. *E. coli* strain BL21 ( $\lambda$ DE3) carrying pVC70108 was used to express the chimaeric protein upon IPTG induction. Amp<sup>r</sup>,  $\beta$  lactamase gene; B, *Bam*HI; A, *Ava*I. b, Schematic arrangement of various domains of anti-Tac(Fv)-PE40. The hybrid protein is shown as a single polypeptide chain. The C-terminus of VH is joined to the N-terminus of VL through a 15 amino-acid linker as shown in a. Details of the construction will be described elsewhere or can be obtained from the authors.

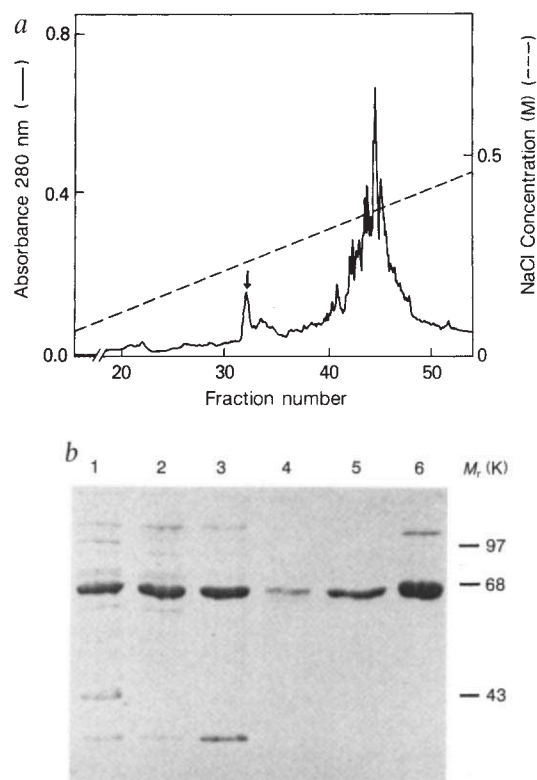


FIG. 2 Purification and characterization of anti-Tac(Fv)-PE40. a, Mono Q column chromatography of renatured soluble anti-Tac(Fv)-PE40. Renatured material was applied on a Mono Q column; proteins were eluted with a NaCl gradient (0-0.5M) and 4 ml fractions were collected. The position of active monomeric anti-Tac(Fv)-PE40 is shown by a vertical arrow. b, SDS-PAGE of samples at various stages of purification. The gel was stained with Coomassie blue. Lane 1, total cell pellet; lane 2, spheroplasts; lane 3, 100,000 g pellet of sonicated spheroplasts; lane 4, pool of fractions (32-33) from the mono Q column; lane 5, pool of peak fractions from the TSK-250 column; lane 6, native PE,  $M_r$  66 K. Molecular weight markers (K) are indicated. METHODS. *E. coli* strain BL21 ( $\lambda$ DE3) carrying plasmid pVC70108 was grown, induced with IPTG, and the cell pellet was processed as described previously<sup>6-9</sup>. Dithiothreitol was omitted from denaturation buffer and renaturation was carried out for 16 h. After renaturation and dialysis, the sample was applied on a Mono Q column (HR 10/10) at 3 ml min<sup>-1</sup>. The column was washed with 40 ml Buffer A (Tris-HCl 20 mM, pH 7.6) and developed with a 200 ml linear gradient (0-0.5 M NaCl). Eluted proteins were monitored at 280 nm. Fractions (4 ml) were collected and tested for cytotoxicity on HUT-102 cells. For SDS-PAGE, samples were boiled with Laemmli sample buffer<sup>18</sup> and electrophoresed on a 10% gel.



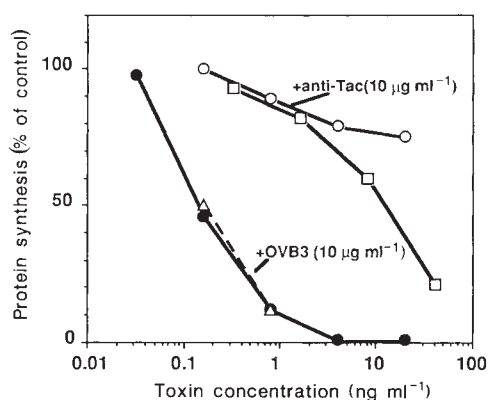


FIG. 3 Cytotoxicity of anti-Tac(Fv)-PE40 and anti-Tac-PE40 on HUT-102 cells expressing IL-2 receptors. Cytotoxicity was determined by measuring protein synthesis in HUT-102 cells after treatment with: (●) anti-Tac(Fv)-PE40; (○) anti-Tac(Fv)-PE40 + 10  $\mu$ g anti-Tac; (△) anti-Tac(Fv)-PE40 + 10  $\mu$ g OVB3; (□) anti-Tac-PE40.

METHODS. HUT-102 cells were washed twice with serum-free medium and plated in RPMI 1640 medium with 5% fetal bovine serum at  $3 \times 10^5$  cells per well in 24-well plates. Various dilutions of recombinant anti-Tac(Fv)-PE40 and chemically conjugated anti-Tac-PE40 were prepared in PBS with 0.2% human serum albumin and added to appropriate wells. After 20 h the cells were labelled with [ $^3$ H]leucine for 90 min and the radioactivity in the TCA precipitate of the cell pellet determined. The results are expressed as % of control with no toxin added. For competition, 10  $\mu$ g of anti-Tac or OVB3 were added to each well just before adding anti-Tac(Fv)-PE40.

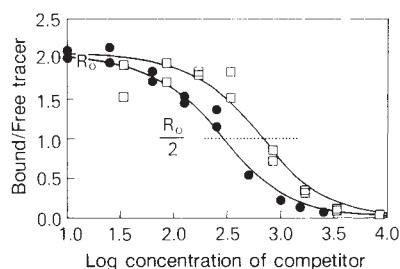


FIG. 4 Competition binding analysis of anti-Tac versus anti-Tac(Fv)-PE40. Competition of anti-Tac(Fv)-PE40 (open squares) and native mouse anti-Tac (solid circles) with [ $^{125}$ I] labelled tracer anti-Tac to bind to Tac antigen on HUT-102 cells is shown. Solid lines are computer generated idealized curves that model binding competition.  $R_0$  is the bound/free ratio for tracer in the absence of competitor, and  $R_0/2$  is the 50% inhibition point for tracer binding, from which a binding affinity of  $3.5 \times 10^9$  M $^{-1}$  for anti-Tac(Fv)-PE40 is calculated, compared with  $9.7 \times 10^9$  M $^{-1}$  for native anti-Tac.

METHODS. [ $^{125}$ I]labelled anti-Tac (2  $\mu$ Ci  $\mu$ g $^{-1}$ ) as tracer was used at 1.5 ng per assay with varying concentrations of competitor and  $4 \times 10^5$  HUT-102 cells as source of Tac antigen in 0.2 ml of binding buffer (RPMI 1640 with 10% fetal bovine serum, 100  $\mu$ g ml $^{-1}$  human IgG, 0.1% sodium azide), and incubated at room temperature with mixing for 2 h. Under these conditions, the concentration of tracer is 50 pM and Tac peptide 500 pM. Free tracer is 10 pM by calculation and satisfies the condition that free tracer be less than  $1/K_a = 100$  pM (using  $10^{10}$  M $^{-1}$  for anti-Tac  $K_a$ ) for the assumptions of the competition analysis<sup>19</sup>. Assays were performed in parallel with a control cold anti-Tac antibody, and curve shifts at the 50% inhibition point of bound/free tracer binding ( $R_0/2$ ) versus log competitor concentration were quantitated. The concentrations were obtained from the antilogs of the abscissa, and the affinity constant  $K_x$  for the construct,  $X$ , derived from the formula<sup>19</sup>  $([X]_{1/2} - [\text{anti-Tac}]_{1/2}) = 1/K_x - 1/K_a$  where  $[ ]_{1/2}$  indicates the concentration of competitor at which tracer binding is  $R_0/2$ . Standard Scatchard plotting of binding data with anti-Tac gave linear graphics and a  $K_a$  of  $9.7 \times 10^9$  M $^{-1}$ , comparable to that obtained by other investigators<sup>16</sup>. The  $K_x$  of  $3.5 \times 10^9$  M $^{-1}$  for anti-Tac(Fv)-PE40 was calculated from the above formula. All concentrations were measured by Bradford protein micro-assay against a standard curve with human IgG (ref. 20). For competition analysis, these concentrations were normalized on the basis of the bindable fraction obtained in separate tests with radiolabelled anti-Tac(Fv)-PE40 (0.44) and radiolabelled anti-Tac (0.8) with excess HUT-102 cells<sup>16</sup> to yield concentrations of bindable protein for the abscissa.

TABLE 1 Cytotoxicity of anti-Tac(Fv)-PE40 on various cell lines

| Cell line | IL-2 receptors per cell |               | ID <sub>50</sub><br>ng ml $^{-1}$ |
|-----------|-------------------------|---------------|-----------------------------------|
|           | Low affinity            | High affinity |                                   |
| HUT-102   | 94,000                  | 3,800         | 0.15                              |
| Cr II.2   | 12,000                  | 350           | 2.7                               |
| CEM       | <20                     | <20           | >1,000                            |
| OVCAR-3   | —                       | —             | >1,000                            |
| KB        | —                       | —             | >1,000                            |
| A431      | —                       | —             | >1,000                            |

Cell lines OVCAR3, KB and A431 were seeded at  $1 \times 10^5$  ml $^{-1}$  in 24-well plates one day before the addition of toxin. HUT-102, Cr II.2 and CEM were washed twice and seeded at  $3 \times 10^5$  ml $^{-1}$  in 24-well plates (see also Fig. 3). RPMI 1640 with 10% fetal bovine serum was used for Cr II.2. Various dilutions of toxin preparations were added, and 20 h later the cells were labelled for 90 min with [ $^3$ H]leucine. The radioactivity in the trichloroacetic acid precipitate of the cells was determined. ID<sub>50</sub> is the concentration of toxin that inhibits protein synthesis by 50% as compared with a control with no toxin added. All the assays were done in duplicate and repeated three times. (Data for the number of low and high-affinity IL-2 receptors on the various cell lines is from the unpublished data of T. Waldmann and ref. 16.)

of both low- and high-affinity IL-2 receptors than HUT 102, was also sensitive to anti-Tac(Fv)-PE40, with an ID<sub>50</sub> of 2.7 ng ml $^{-1}$  (Table 1). Furthermore, several human cell lines without IL-2 receptors, including the T-cell leukaemia line CEM, as well as carcinoma cell lines A431, KB and OVCAR-3 (ref. 14), were not affected by anti-Tac(Fv)-PE40, even at 1  $\mu$ g ml $^{-1}$  (Table 1).

Previously we reported that anti-Tac chemically conjugated to PE- or PE40-killed HUT-102 cells<sup>14,15</sup>. When thioether conjugates are made, anti-Tac-PE had an ID<sub>50</sub> of 1.2 ng ml $^{-1}$ , and anti-Tac-PE40 similarly prepared had an ID<sub>50</sub> of 13 ng ml $^{-1}$ . As anti-Tac(Fv)-PE40 (65 K) is about 30% smaller by weight than anti-Tac-PE (216 K), the chimaeric toxin is on a molar basis several times more active than anti-Tac-PE and considerably more active than anti-Tac-PE40. Anti-Tac-PE is a heterogeneous chemical conjugate in which the two molecules are connected by a thioether bond and different lysines in PE and anti-Tac are used in the conjugation reaction. In anti-Tac-PE40, the attachment appears to be mainly through the lysine residues in domain III of PE40 and this reduces the activity of the PE conjugates<sup>14</sup>. IL-2-PE40 is another chimaeric molecule, which was constructed by fusing a complementary DNA for human IL-2 to PE40 sequences<sup>7</sup>. IL-2-PE40 is slightly less cytotoxic to HUT-102 cells than anti-Tac(Fv)-PE40, with an ID<sub>50</sub> of 1–5 ng ml $^{-1}$ .

Competition binding studies showed an affinity of  $3.5 \times 10^9$  M $^{-1}$  for anti-Tac(Fv)-PE40, ~3-fold lower than that of anti-Tac, measured at  $9.7 \times 10^9$  M $^{-1}$  (Fig. 4). This can be compared with a fourfold loss of affinity of a Fv construct versus Fab fragment of anti-bovine growth hormone<sup>3</sup> and a sixfold loss for the Fv construct versus an Fab fragment of anti-digoxin<sup>4</sup>. The affinity of Fv binding for p55 seems to be preserved to a greater extent than similar preparations of anti-bovine growth hormone and anti-digoxin antibodies.

In summary, we have created an active immunotoxin in *E. coli* by the fusion of cDNAs encoding the anti-Tac variable regions with a fragment of DNA encoding a modified form of *Pseudomonas* exotoxin. Using this approach it should now be possible to create active recombinant immunotoxins with other antibodies. □

Received 16 February; accepted 12 April 1989.

- Pastan, I., Willingham, M. C. & FitzGerald, D. J. *Cell* **47**, 641–648 (1986).
- Vitetta, E. S., Fulton, R. J., May, R. D., Till, M. & Uhr, J. W. *Science* **238**, 1098–1104 (1987).
- Bird, R. E. *et al. Science* **242**, 423–426 (1988).
- Houston, J. S. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 5879–5883 (1988).
- Uchiyama, T., Broder, S. & Waldmann, T. A. *J. Immuno.* **126**, 1393–1397 (1981).
- Chaudhary, V. K., FitzGerald, D. J., Adhya, S. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4538–4542 (1987).

7. Lorberboum-Galski, H., Fitzgerald, D., Chaudhary, V., Adhya, S. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1922-1926 (1988).
8. Chaudhary, V. K. *et al. Nature* **335**, 369-372 (1988).
9. Siegall, C. B., Chaudhary, V. K., Fitzgerald, D. J. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9738-9742 (1988).
10. Bacha, P., *et al. Proc. natn. Acad. Sci. U.S.A.* **167**, 612-622 (1988).
11. Studier, F. W. & Moffat, B. A. *J. molec. Biol.* **189**, 113-130 (1986).
12. Willingham, M. C., Fitzgerald, D. J., & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2474-2478 (1987).
13. Salahuddin, S. Z. *et al. Virology* **129**, 51-64 (1983).
14. Kondo, T., Fitzgerald, D., Chaudhary, V., Adhya, S., & Pastan, I. *J. biol. Chem.* **263**, 9470-9475 (1988).
15. Fitzgerald, D. J. P., Waldmann, T. A., Willingham, M. C. & Pastan, I. *J. clin. Invest.* **74**, 966-971 (1984).
16. Robb, R. J., Green, W. G., & Rusk, C. M. *J. exp. Med.* **160**, 1126-1146, (1984).
17. Hwang, J., Fitzgerald, D. J. P., Adhya, S. & Pastan, I. *Cell* **48**, 129-136 (1987).
18. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
19. Berzofsky, J. A. & Schechter, A. N. *Molec. Immun.* **18**, 751-763, (1981).
20. Bradford, M. M. *Analyt. Biochem.* **72**, 248-254, (1976).

ACKNOWLEDGEMENTS. We thank B. Lovelace for technical assistance, S. Neal for photographic work, A. Gaddis and J. Evans for typing the manuscript and Dr C. Klee for performing the amino-acid sequence analysis.

## Vesicular transport between the endoplasmic reticulum and the Golgi stack requires the NEM-sensitive fusion protein

Con J. M. Beckers\*, Marc R. Block†‡, Benjamin S. Glick†‡, James E. Rothman|| & William E. Balch\*

\* Department of Molecular Biology, Scripps Clinic and Research Foundation, La Jolla, California 92037, USA

† Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA

|| Department of Biology, Lewis Thomas Laboratory, Princeton University, Princeton, New Jersey 08544, USA

AN *N*-ethylmaleimide-sensitive fusion protein (NSF) has been purified on the basis of its ability to catalyse vesicular transport within the Golgi stack. We report here that this same protein is required for transport from the endoplasmic reticulum to the Golgi stack in semi-intact cells. This transport process is inhibited by a monoclonal antibody against NSF. Furthermore, pretreatment of semi-intact cells with *N*-ethylmaleimide, a sulphhydryl alkylating reagent, inhibits transport. Addition of highly purified NSF largely restores transport from endoplasmic reticulum to Golgi. These results suggest that NSF is a general component of the transport machinery required for membrane fusion at multiple stages of the secretory pathway.

The *SEC18* gene of yeast encodes an NSF activity that will function in place of animal cell NSF with animal cell Golgi membranes<sup>1</sup>. As mutants in the *SEC18* gene are defective in transport from the endoplasmic reticulum (ER) to the Golgi in yeast<sup>2</sup> we wondered whether NSF is also required for this transport step in animal cells, in addition to its established role in promoting fusion within the confines of the Golgi stack. To address this issue we have examined whether NSF is needed for transport of the vesicular stomatitis virus (VSV)-encoded glycoprotein (G protein) between the ER and *cis* Golgi compartment in an *in vitro* system<sup>3</sup> in which semi-intact cells prepared from VSV-infected Chinese hamster ovary (CHO) cells are incubated in the presence of cytosol and ATP.

A complete incubation mix containing semi-intact cells, cytosol and ATP was treated with *N*-ethylmaleimide (NEM) on ice for 15 min (conditions which inactivate NSF<sup>4</sup>) before incubation. Subsequent transport (during an incubation at 30 °C for 90 min) was inhibited by more than 90 per cent (Fig. 1, lane iv). To determine whether the NEM-sensitive factor required

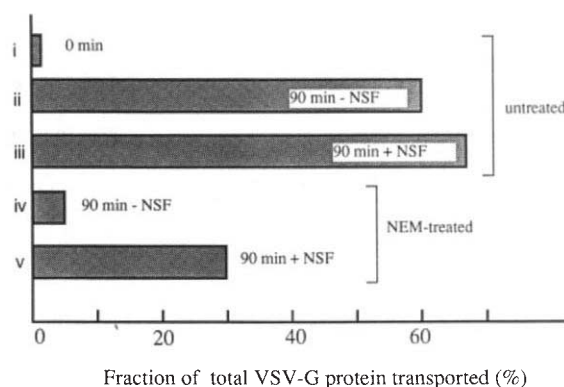


FIG. 1 NSF is required for transport from ER to Golgi. Assay conditions were as previously described<sup>3,6</sup>. Lanes i-iii, semi-intact cells, cytosol and ATP were incubated at 30 °C for the indicated time in the absence (lane ii) or presence (lane iii) of 0.16 µg of NSF (purified from CHO cells as described<sup>5</sup> in a final volume of 40 µl. Lanes iv and v, semi-intact cells and cytosol were pretreated with 1 mM NEM for 15 min at 0 °C. Subsequently, glutathione was added to a final concentration of 2 mM to quench unreacted NEM. NEM-treated membranes and cytosol were incubated in the presence of ATP at 30 °C for 90 min in the absence (lane iv) or presence (lane v) of 0.16 µg NSF. The fraction of total VSV-G transported was measured as the fraction of G protein in the Man<sub>5</sub> form as described previously<sup>3,6</sup>.

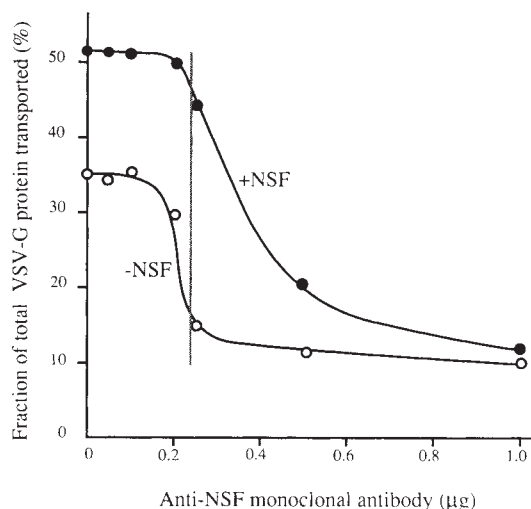


FIG. 2 Anti-NSF antibody inhibits the transport of VSV-G protein to the *cis* Golgi compartment. Assay conditions were as previously described<sup>3,6</sup>. The indicated amount of anti-NSF antibody was added to a 40 µl complete cocktail containing semi-intact cells, cytosol and ATP for 15 min on ice. Subsequently, the cocktail was transferred to 30 °C and incubated for 90 min in the absence (open circles) or presence of 0.16 µg purified NSF. The anti-NSF antibody (4A6) is an IgM whose properties have been described previously<sup>5</sup>, and was purified from ascites fluid as described<sup>5</sup>.

for delivery of VSV-G protein from the ER to the *cis* Golgi compartment could be replaced by NSF, NEM-treated semi-intact cells and NEM-treated cytosol were incubated together at 30 °C in the presence or absence of NSF purified from CHO cells as described<sup>5</sup>. Addition of NSF stimulated transport about 10-fold over background (Fig. 1, compare lanes iv and v) restoring activity to a level which was nearly 50 per cent of that of an untreated incubation that had been supplemented with NSF (Fig. 1, lane iii). In this instance, NSF only marginally (10%) stimulated an incubation that had not been pretreated with NEM (Fig. 1, compare lanes ii and iii). But some preparations of semi-intact cells (apparently more deficient in NSF) are stimulated by up to 30 per cent (see Fig. 2). These results strongly

‡ Present addresses: Laboratoire de Physiologie Cellulaire, Université Joseph Fourier, B.P. 53X, 38041 Grenoble Cedex, France (M.R.B.); Biocenter, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland (B.S.G.).

suggest that NSF can efficiently substitute for a factor(s) inactivated by the pretreatment with NEM and required for the transport of VSV-G protein from the ER to the *cis* Golgi compartment.

This restorative activity of NSF could have either of two interpretations. One possibility is that the NEM-sensitive factor required for ER to Golgi transport is a protein unrelated structurally to NSF, but one whose function can be replaced by NSF. Alternatively, the NEM-sensitive protein needed for ER to Golgi transport may be closely related, if not identical, to NSF. To distinguish these possibilities, semi-intact cells and cytosol were preincubated for 15 min on ice in the presence of various amounts of an IgM monoclonal antibody specific for NSF (anti-NSF 4A6 has been previously demonstrated to inhibit potentially intra-Golgi transport<sup>5</sup>). Then, the assay mix was transferred to 30 °C and incubated for 90 min in the presence of ATP. As shown in Fig. 2 (open circles), the anti-NSF monoclonal antibody strongly inhibited transport. Inhibition of transport by anti-NSF could be readily overcome by the addition of purified NSF (Fig. 2, closed circles). The addition of 0.16 µg purified NSF to the incubation required an additional 0.55 µg monoclonal antibody to obtain maximum inhibition of transport (Fig. 2, closed circles), corresponding to nearly equimolar amounts of IgM pentamer per mole of NSF tetramer bound. These results suggest that the NEM-sensitive component required for ER to Golgi transport and the intra-Golgi transport fusion protein NSF are highly related, if not identical, proteins.

Does NSF fulfill the function of a fusion protein in transport from ER to Golgi, as it does for transport of protein between

Golgi compartments? Recent evidence supports this notion<sup>6</sup>. Transport of protein from the ER to the *cis* Golgi can be inhibited by the addition of GTP-γ-S or by EGTA. Addition of GTP-γ-S results in the accumulation of VSV-G protein in a post-ER transport intermediate<sup>6</sup>; incubation in the presence of EGTA results in accumulation of VSV-G protein in a transport intermediate whose consumption upon adding 0.1 µM free Ca<sup>2+</sup> is resistant to GTP-γ-S (Fig. 3b, lane vii)<sup>6</sup>. This indicates that transport from the intermediate requiring GTP hydrolysis precedes the Ca<sup>2+</sup>-requiring step. To test whether this post-ER transport intermediate requires NSF for fusion to the *cis* Golgi compartment, VSV-G protein was accumulated in the EGTA-sensitive intermediate for 60 min. When NEM was added before reincubation in the presence of 0.1 µM Ca<sup>2+</sup> no transport was observed, suggesting that the NEM-sensitive component was still required for compartment. To test whether this NEM-sensitive component was NSF, the EGTA-sensitive intermediate was incubated in the presence of Ca<sup>2+</sup> and the monoclonal antibody to NSF. As shown in Fig. 3, anti-NSF antibody inhibited transport from the intermediate (Fig. 3b, lane ix) to the same extent as that observed from the ER (Fig. 3a, lane iv) suggesting that NSF is required at a late step preceding vesicle fusion to the *cis* Golgi compartment.

In summary, NSF seems to function as a fusion protein required for delivery of cargo protein to all compartments of the Golgi stack, independent of vesicle origin. This suggests that the fusion of a carrier vesicle to a Golgi cisterna may have a common biochemical foundation, whether the vesicle comes from the ER, from within the Golgi stack, or possibly from other sites. □

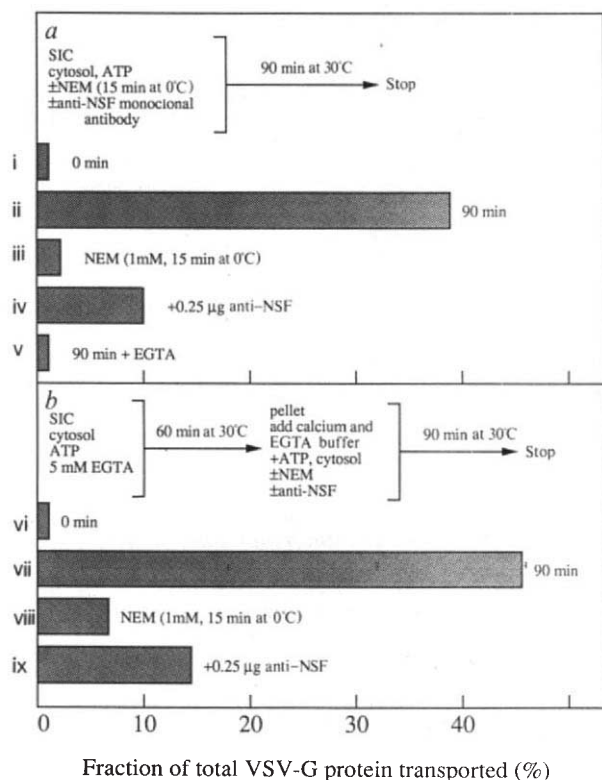


FIG. 3 Anti-NSF antibody inhibits a late step in transport to the *cis* Golgi compartment. Assay conditions were as previously described<sup>3,6</sup>. NEM treatment (lanes iii and viii), was as described in Fig. 1. Where indicated (lanes iv and ix) 0.25 µg anti-NSF antibody was added to 40 µl a complete cocktail containing semi-intact cells (SIC), cytosol and ATP for 15 min on ice. Subsequently, membranes containing VSV-G in the ER (a) or the EGTA-sensitive intermediate (b) were incubated for 90 min at 30 °C. The amount of transport observed after incubation for 60 min in the presence of EGTA is shown in lane vi (0 min). After incubation in the presence of EGTA for 60 min (b), free Ca<sup>2+</sup> was adjusted to 0.1 µM by addition of an EGTA/Ca<sup>2+</sup> buffer as described previously<sup>6</sup>.

Received 23 March; accepted 21 April 1989

1. Wilson, D. W. *et al. Nature* **339**, 335–359 (1989).
2. Novick, P. & Schekman, R. *Cell* **21**, 205–215 (1980).
3. Beckers, C. J. M., Keller, D. S. & Balch, W. E. *Cell* **50**, 523–534 (1987).
4. Glick, B. S. & Rothman, J. E. *Nature* **326**, 309–312 (1987).
5. Block, M. R., Glick, B. S., Wilcox, C. A., Weiland, F. T. & Rothman, J. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7852–7856 (1988).
6. Beckers, C. J. M. & Balch, W. E. *J. Cell Biol.* **108**, 1245–1256 (1989).

ACKNOWLEDGEMENTS. This work was supported by the NIH (W.E.B. and J.E.R.) and the Harold and Leila Mathers Foundation (W.E.B.). M.B. was supported by a California American Cancer Society Senior Fellowship.

## Vesicle fusion following receptor-mediated endocytosis requires a protein active in Golgi transport

Ruben Diaz\*, Luis S. Mayorga\*, Peggy J. Weidman†, James E. Rothman† & Philip D. Stahl\*

\* Department of Cell Biology and Physiology, Washington University School of Medicine, St Louis, Missouri 63110, USA

† Department of Biology, Lewis Thomas Laboratory, Princeton University, Princeton, New Jersey 08544, USA

IN reconstitution studies *N*-ethylmaleimide, a sulphydryl alkylating reagent, inhibits both fusion of endocytic vesicles<sup>1–3</sup> and vesicular transport in the Golgi apparatus<sup>4</sup>. We show here that the same *N*-ethylmaleimide-sensitive factor that catalyses the vesicle-mediated transport within Golgi stacks is also required for endocytic vesicle fusion. Thus, it is likely that a common mechanism for vesicle fusion exists for both the secretory and endocytic pathways of eukaryotic cells.

A number of cell-free assays have been described recently to study endocytic vesicle fusion<sup>1–3,5–7</sup>. We use an assay<sup>1</sup> based on two probes that are internalized through cell surface receptors



and form a measurable complex when they encounter each other in the same luminal compartment. These probes are the dinitrophenol (DNP) derivative of  $\beta$ -glucuronidase, a ligand for the macrophage mannose receptor, and a mannose derivative of anti-DNP mouse monoclonal IgG, a ligand for the same mannose receptor. Each ligand is internalized into separate populations of macrophages, which are then homogenized. The post-nuclear supernatants are incubated together *in vitro*. Fusion of endocytic vesicles containing enzyme and antibody allows the formation of an immune complex which can be quantitated as a measure of fusion. Endocytic vesicle fusion in this<sup>1</sup> and other assays<sup>2,3,5</sup> requires ATP, cytosol and membrane proteins. In addition, treatment of the reaction mixture with *N*-ethylmaleimide (NEM) (1 mM for 60 min at 4 °C in our assay) results in complete inhibition of the fusion reaction. Treatment of either vesicle fraction or a high-speed supernatant from a cell homogenate separately with NEM does not result in full inhibition, suggesting that the factor(s) affected by NEM could be present in both soluble and membrane-associated form<sup>1</sup>.

Preparations containing both vesicles and cytosol were first treated with NEM and the residual presence of the sulphydryl agent was quenched by excess dithiothreitol. This mixture was then incubated under fusogenic conditions in the presence of fresh cytosol. Addition of cytosol fractions from various cell lines and tissue sources restored fusion. Among those tested,

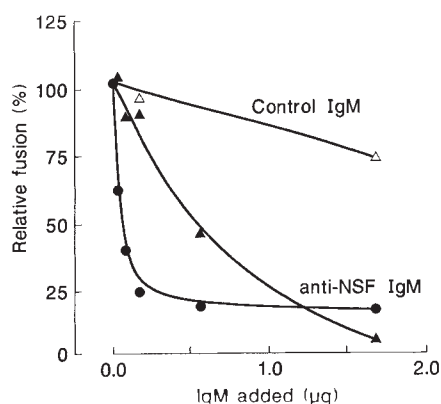


FIG. 1 Inhibition by anti-NSF antibody of cytosol-mediated restoration of endocytic vesicle fusion. Post-nuclear supernatants containing the mannose derivative of anti-DNP IgG (Man-IgG) and the dinitrophenol derivative of  $\beta$ -glucuronidase (DNP- $\beta$ -glucuronidase) in 5-min endosomes were incubated with *N*-ethylmaleimide (NEM) (1 mM for 60 min at 4 °C). Excess NEM was quenched with 2 mM dithiothreitol (DTT). Post-nuclear supernatants were then mixed at 4 °C in fusion buffer (250 mM sucrose, 0.5 mM EGTA, 1 mM DTT, 1.5 mM  $MgCl_2$ , 250  $\mu g\ ml^{-1}$  DNP-derivative of bovine serum albumin (DNP-BSA), 50 mM KCl, 10  $\mu M$  palmitoyl coenzyme A, 1 mM ATP, 8 mM creatine phosphate, 31 units per ml creatine phosphokinase, 20 mM HEPES-KOH, pH 7). The assay consisted in mixing 10  $\mu l$  of the preparation ( $\sim 15\ \mu g$  cytosol protein and 5  $\mu g$  membrane protein) with 2.5  $\mu l$  fusion buffer containing 5  $\mu g$  CHO ( $\Delta$ ,  $\triangle$ ) or J774 ( $\bullet$ ) cytosol that had been pre-incubated for 60 min at 4 °C with increasing amounts of either control IgM or 4A6 IgM (anti-NSF IgM). Incubations were carried out for 40 min at 37 °C. Fusion was assessed by formation of complex between the antibody and the enzyme. The complex was immunoprecipitated using Staph A coated with rabbit anti-mouse antibody in a buffer containing detergent and DNP-BSA and quantified by measuring  $\beta$ -glucuronidase activity with a fluorescent assay. Values are reported as percentage of the total recovered fusion with 5  $\mu g$  CHO or J774 cytosol that was not pre-incubated with antibody. The signal obtained in the absence of either J774 or CHO cytosol was subtracted from all values. Post-nuclear supernatants were generated as previously described<sup>1</sup>, from two populations of J774 E clone macrophages (a mannose receptor positive cell line) incubated with either 10  $\mu g\ ml^{-1}$  Man-IgG or 20  $\mu g\ ml^{-1}$  DNP- $\beta$ -glucuronidase respectively at 37 °C for 5 min. Cytosols were the high-speed supernatants (150,000g, 60 min) of cell homogenates (10% suspensions) of CHO and J774 cell lines. Cytosols were stored frozen until use.

both J774 macrophage and chinese hamster ovary (CHO) cytosol were effective in completely recovering fusion activity following NEM inactivation. Treatment of these fractions with NEM before their addition to the assay abolished the restorative effect. There was no substantial increase in fusogenic activity when cytosol fractions were added to post-nuclear supernatants which had not been treated with NEM. This suggested that any other soluble proteins required for endosome fusion were already present at saturating levels within the post-nuclear supernatants, so that only the NEM-sensitive restorative activity of the cytosol was being measured.

We tested the effect of a monoclonal IgM antibody<sup>8</sup> that inactivates the NEM-sensitive factor (NSF)<sup>4</sup> on the restorative activity of CHO cytosol. NSF is a tetramer made up of four polypeptide chains, each with a relative molecular mass of 76,000. It is required for the fusion of transport vesicles to Golgi membranes<sup>8-10</sup>. The NEM-sensitivity of this protein and its established role in vesicular transport make it a good candidate for a role in vesicle fusion during endocytosis. Addition of 5  $\mu g$  J774 macrophage or CHO cytosol (per 12.5  $\mu l$  assay volume) restored 40% and 30% of the total fusogenic activity, respectively. When CHO cytosol was pre-incubated with anti-NSF IgM before its addition to the assay, the antibody inhibited the recovery of fusion by as much as 90% (Fig. 1). J774 macrophage cytosol proved to be even more sensitive to immune inactivation than CHO cytosol. Control IgM antibodies were not inhibitory (Fig. 1).

To establish further whether NSF plays a role in endocytic vesicle fusion, pure CHO NSF was added back to NEM-treated post-nuclear supernatants to assess whether it could restore fusogenic activity (Fig. 2). There was an increase in fusogenic activity with increasing amounts of NSF, reaching a plateau at around 20 ng (per 12.5  $\mu l$ ). The restorative activity of 54 ng NSF amounted to 65 per cent of the activity obtained when 5  $\mu g$

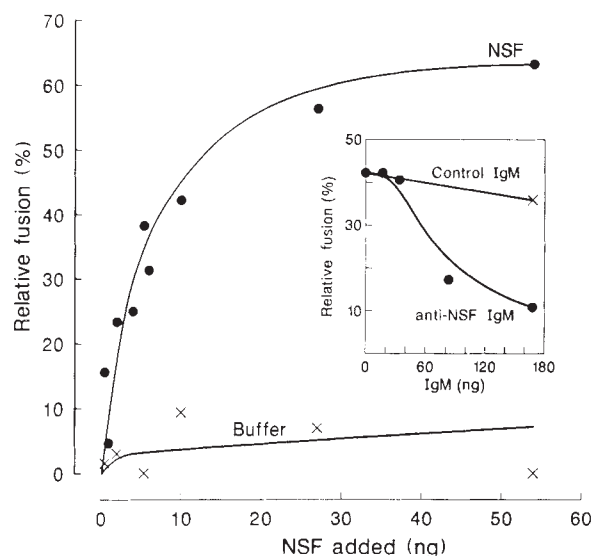


FIG. 2 Restoration of endocytic vesicle fusion activity by purified NSF. Post-nuclear supernatants containing Man-IgG and DNP- $\beta$ -glucuronidase in 5 min endosomes were inactivated with NEM and mixed in fusion buffer as described in Fig. 1. Increasing amounts of pure CHO NSF (maintained at 27  $\mu g\ ml^{-1}$  in 100 mM KCl, 2 mM  $MgCl_2$ , 1 mM DTT, 1% polyethylene glycol, 10% glycerol, 250  $\mu g\ ml^{-1}$  soybean trypsin inhibitor, 500  $\mu M$  ATP, 20 mM HEPES, pH 8) were added to NEM-inactivated post-nuclear supernatants. As a control, only the buffer containing NSF was added (buffer). NSF-dependent restoration of fusogenic activity is expressed as a percentage of the fusion activity recovered with 5  $\mu g$  CHO cytosol. Incubations were carried out for 40 min at 37 °C. The inset shows the inhibition by anti-NSF IgM of the recovery of fusion with 10 ng NSF. 10 ng of NSF were incubated for 30 min at 4 °C with either 4A6 anti-NSF IgM or control IgM before being added to the assay, and the recovery of fusogenic activity was measured.

CHO cytosol were added back. The restorative activity of the purified NSF was inhibited by the monoclonal antibody (Fig. 2, inset). The molar ratio of pentameric antibody to NSF tetramers required to obtain half-maximal inhibition of the recovered fusogenic activity ( $\sim 1.5$ ) resembles the ratio reported to inhibit NSF activity in Golgi transport ( $\sim 0.96$ )<sup>8</sup>.

NSF is highly sensitive to heat inactivation in the absence of ATP<sup>8</sup>. We observed that the NEM-sensitive restorative activity for endocytosis of our cytosol preparations from both CHO and J774 cell lines remained stable at 37 °C as long as they had not been gel-filtered to remove residual adenine nucleotides. Once the cytosols had been gel-filtered, the restorative NEM-sensitive fusogenic activity was lost after only a few minutes of incubation at 37 °C (data not shown). ATP (1 mM) stabilizes the factor, as has been observed for NSF within the context of Golgi transport<sup>8</sup>.

Our results indicate that the fusion among early endocytic vesicles which contain ligand internalized by receptor-mediated endocytosis requires NSF. Thus, our experiments clearly extend the scope of the activities of NSF beyond the secretory pathway. Exactly which step(s) and endocytic vesicles use NSF, or whether they all do, remains to be determined. Anti-NSF antibody has also an inhibitory effect on reconstitution assays where endocytic vesicles fuse with plasma membrane-derived vesicles<sup>6</sup> or with protease containing compartments<sup>7</sup> (data not shown).

We conclude that NSF plays an important part in fusion of endocytic vesicles during early stages of receptor-mediated endocytosis. NSF appears to require a set of both soluble and membrane proteins to associate with Golgi membranes<sup>11</sup>. It has

not been determined whether these other components are also required for the activity of NSF in the endocytic pathway. NSF, nevertheless, appears to be a component of a machinery available to mammalian and other eukaryotic cells to catalyse the fusion of intracellular membranes. In addition to its requirement for Golgi transport, there is now evidence that vesicle transport from endoplasmic reticulum to Golgi also requires NSF<sup>12</sup>. The ability of NSF to catalyse vesicle fusion in both secretory and endocytic pathways implies that the factor itself may not constitute a recognition signal for the regulation of vesicle fusion specificity. Other recognition signals must be present on the cytoplasmic surface of these vesicles that determine the fate of the transport vesicle. □

Received 23 March; accepted 28 April 1989.

1. Diaz, R., Mayorga, L. & Stahl, P. *J. biol. Chem.* **263**, 6093–6100 (1988).
2. Braell, W.A. *Proc. Natn. Acad. Sci. U.S.A.* **84**, 1137–1141 (1987).
3. Woodman, P.G. & Warren, G. *Eur. J. Biochem.* **173**, 101–108 (1988).
4. Glick, B.S. & Rothman, J.E. *Nature* **326**, 309–312 (1987).
5. Gruenberg, J.E. & Howell, K.E. *EMBO J.* **5**, 3091–3101 (1986).
6. Mayorga, L.S., Diaz, R. & Stahl, P.D. *J. biol. Chem.* **263**, 17213–17216 (1988).
7. Mayorga, L.S., Diaz, R. & Stahl, P.D. *J. biol. Chem.* **264**, 5392–5399 (1989).
8. Block, M.R., Glick, B.S., Wilcox, C.A., Wieland, F.T. & Rothman, J.E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7852–7856 (1988).
9. Malhotra, V., Orci, L., Glick, B.S., Block, M.R. & Rothman, J.E. *Cell* **54**, 221–227 (1988).
10. Wilson, D.W. *et al. Nature* **339**, 355–359 (1989).
11. Weidman, P.J., Melancon, P., Block, M.R. & Rothman, J.E. *J. Cell Biol.* **108**, 1589–1596 (1988).
12. Beckers, C.J.M., Block, M.R., Glick, B.S., Rothman, J.E. & Balch, W.E. *Nature* **339**, 397–398 (1989).

ACKNOWLEDGEMENTS. We thank Lia E.B. Mayorga for technical assistance. The work was supported, in part, by the NIH. L.M. is a fellow of the National Research Council of Argentina. P.W. is a fellow of the NIH.

## CORRIGENDUM

### *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts

Clifford W. Morden & Susan S. Golden

*Nature* **337**, 382–385 (1989)

THE phylogenetic tree shown in Fig. 3 of this letter was generated using a data set in which numeric coding errors have been found. These were detected by using a new version of PAUP<sup>1</sup> which interprets the single-letter amino acid code directly and does not require coding. The uncoded amino acid data set was analysed as previously reported, using the branch and bound option of PAUP<sup>1</sup>. In the previous analysis many equally parsimonious trees were obtained when different treatments were used to accommodate the seven-amino-acid domain missing in *Prochlorothrix hollandica* and green chloroplast-containing taxa. In the current analysis only two phylogenetic tree topologies were obtained, corresponding to the published tree shown in Fig. 3 and a variation described in the text, in which *P. hollandica* was associated in a lineage including the *Synechococcus* sequences. The only variations observed were slight changes in branch lengths among the different treatments.

When the seven-amino-acid domain missing in some taxa was excluded from the analysis, or when its presence or absence

was treated as a single binary character, a unique phylogenetic tree was generated associating the *P. hollandica* and *Synechococcus* sequences. Two methods were used to consider presence or absence of the domain as equivalent to two amino acid changes. If each character in the seven-amino-acid domain was weighted two-sevenths, one tree was obtained associating *P. hollandica* with *Synechococcus*. However, when the seven amino acids were treated as 'missing data' from the taxa that lack those residues, and an additional character having a weight of two was added to designate presence or absence of that domain, two trees were generated: one as described above and one like that shown in Fig. 3. Any additional significance ascribed to the seven-amino-acid domain, from treating it as the equivalent of three to seven amino acid changes, produced only the phylogenetic tree shown in Fig. 3.

This analysis showed that the distribution of the seven-amino-acid domain among taxa is the chief characteristic of the *psbA* genes, suggesting that *P. hollandica* is part of a lineage that led to chloroplasts after a divergence from the cyanobacteria. Phylogenetic analysis of the amino-acid sequences supports this relationship only if the presence or absence of residues in this domain is considered to be more significant than one or two amino-acid substitutions. □

1. Swafford, D. L. PAUP version 3.0 (β-test version, unpublished)

# Replicable RNA reporters

F. R. Kramer and P. M. Lizardi

Elusive infectious agents such as HIV-1 can be detected by nucleic acid hybridization using special probes that are amplified by the enzyme Q-beta replicase as much as a thousand million-fold.

It is imperative that very sensitive assays be developed for the detection of pathogenic retroviruses such as HIV-1. As few as one in 100,000 peripheral blood mononuclear cells may be infected in an asymptomatic individual<sup>1</sup>, yet blood from that person readily infects others. Suitable assays would employ oligonucleotide probes, because they form extremely stable and specific hybrids with the nucleic acids of HIV-1. The sensitivity of these assays depends on their ability to detect the probes once they are bound to their targets. The classic detection strategy is to attach reporter groups to the probes, such as fluorescent molecules or radioactive phosphate groups.

Alternatively, enzymes such as peroxidase or alkaline phosphatase can be linked to the probes and then incubated with a colourless substrate to generate a large amount of coloured product<sup>2</sup>. But the practical detection limit of these schemes is about 200,000 target molecules, which is insufficient to detect the 6,000 molecules of HIV-1 mRNA present in a sample containing a single infected cell.

## Amplifiable probes

An attractive strategy for detecting rare targets is to link each probe to replicatable reporter that can be amplified exponentially after hybridization to reveal the presence of the probe<sup>3</sup>. Our group has developed a novel implementation of this approach, in which a probe sequence is embedded within the sequence of replicatable RNA<sup>4</sup>. The resulting recombinant RNAs hybridize to target sequences as do ordinary probes, and the probe-target hybrids may then be separated from non-hybridized probes, as in an ordinary hybridization assay.

What makes the recombinant-RNA probes unique is that they can then be amplified exponentially by incubation with the RNA-directed RNA polymerase, Q-beta replicase<sup>5</sup>. As many as one thousand million copies of each replicatable probe can be synthesized in a single 30-minute reaction<sup>4</sup>. The extreme specificity of Q-beta replicase for its own template RNA<sup>6</sup> assures that only the replicatable probes are amplified; the large number of copies that are synthesized can easily be quantified by the fluorescence of an intercalating dye, such as ethidium bromide. Because as little as a single molecule of replicatable RNA can initiate exponential amplification<sup>7</sup>, this approach offers the prospect of developing

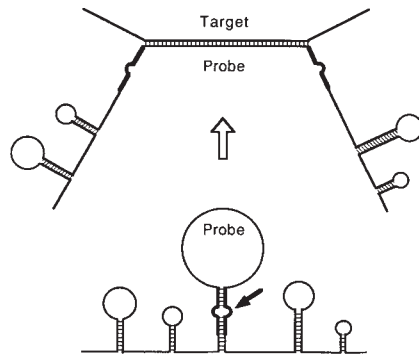


FIG. 1 Replicable RNA containing a molecular switch. Unbound probe forms a cleavage site for ribonuclease III (bold arrow).

extremely sensitive assays.

Two technical advances led to the realization of this amplification scheme: the discovery that oligoribonucleotides can be inserted within the sequence of a small, naturally occurring template for Q-beta replicase, MDV-1 RNA, without interfering with its replicatability<sup>8</sup>, and the availability of a plasmid containing an MDV-1 cDNA sequence that serves as a template for the synthesis of MDV-1 (+) RNA when it is incubated *in vitro* with T7 RNA polymerase<sup>9</sup>. The insertion of probe sequences within the plasmid's MDV-1 sequence creates templates that are used for the synthesis of replicatable probes<sup>4</sup>.

Recombinant RNAs containing an inserted HIV-1 probe<sup>10-12</sup> sequence have been used to explore the characteristics of assays that utilize exponential amplification<sup>13</sup>. Replicable HIV-1 probes efficiently hybridize to target molecules in solutions of the chaotropic salt, guanidine thiocyanate<sup>14</sup>. The resulting probe-target hybrids can be separated from nonhybridized probes by reversible target capture on paramagnetic particles<sup>15</sup>. The replicatable probes are then freed from their targets and used as templates for replication by Q-beta replicase. Kinetic analysis of replication indicates that the number of RNA molecules in a reaction doubles at regular intervals until it equals the number of replicase molecules. The fewer the number of probes that initiate a reaction, the longer it takes to reach this equivalence point. After equivalence is reached, the number of RNA molecules increases linearly. There are two different ways to interpret the amplification results. If the reactions are monitored during the exponential phase, then there is an inverse linear relationship between the logarithm of the number of targets present in a sam-

ple and the time it takes to synthesize a particular, though arbitrary, amount of amplified RNA<sup>15</sup>. Alternatively, if the reactions are observed later, during the linear phase, then there is a direct linear relationship between the logarithm of the number of targets present in a sample and the amount of RNA synthesized at a particular, though arbitrary, time<sup>4</sup>. In either case, the logarithmic nature of the relationship permits a quantitative assessment of the number of targets in a sample, over an unusually wide range of target concentrations.

## Molecular switches

The limiting factor determining the minimum number of targets that can be detected using probe amplification is the persistence of nonhybridized probes, despite extensive washing of the probe-target hybrids<sup>15</sup>. As a result, the current limit of detection is about 10,000 target molecules. To reduce this source of background, recombinant RNAs have been constructed that contain a 'molecular switch', which is a region of the probe molecule that changes its conformation when the probe hybridizes to its target. For example, a molecular switch is created when a probe sequence is embedded between two shorter complementary sequences that pair with each other to form a hairpin stem. When the probe sequence hybridizes to its target, the greater length and relative bending resistance of the resulting duplex<sup>16</sup> forces the sequences in the hairpin stem into a single-stranded conformation.

Figure 1 illustrates how a molecular switch works. In this example, the sequences selected for the molecular switch form a double-stranded recognition site for ribonuclease III (ref. 17). When the probe sequence is hybridized to its target, the complementary sequences are forced apart, eliminating the ribonuclease III binding site. Incubation of the probe-target hybrids with ribonuclease III, prior to amplification, results in the selective destruction of the nonhybridized probes.

## Target-dependent replication

Alternatively, the background due to the persistence of nonhybridized probes can be eliminated by using probes that are not replicatable. In this scheme, a series of enzymatic steps is used to synthesize replicatable RNA reporters. Figure 2 illustrates this scheme and is referred to below. Two different single-stranded



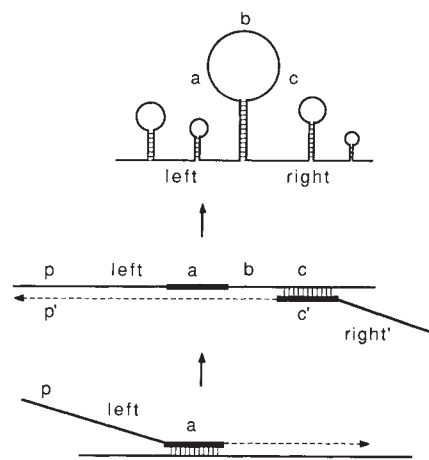


FIG. 2 Target-dependent generation of replicable reporter RNAs. Because the hybridization of two different probes is required for the synthesis of a functional promoter for this artificial gene, signal generation is dependent on the presence of the target.

DNA probes are used. The first DNA (lower diagram) contains a T7 promoter sequence (p), the left half of the MDV-1 sequence, and a probe sequence (a). After this probe is hybridized to its target (a'), it serves as a primer for its own extension<sup>18</sup> by incubation with an appropriate DNA polymerase. The extended sequence includes a copy of the target region (dashed line).

The extended first DNA is then released from its target and hybridized to a second DNA (middle diagram) that contains the right half of the MDV-1 sequence, and a probe sequence (c') that is complementary to the extended sequence of the first DNA. The second DNA then serves as a primer for its own extension by incubation with DNA polymerase. Only if both probes hybridize to their intended targets will a functional double-stranded promoters (p and p') be generated. Incubation of these DNA complexes with T7 RNA polymerase<sup>19</sup> results in the synthesis of replicable recombinant RNAs (upper diagram).

The transcripts are then amplified exponentially by incubation with Q-beta replicase. Because nonhybridized DNA probes are not replicable, signal generation is target-dependent, and a very small number of targets can be detected. An additional feature of this scheme is that the amplified RNAs contain a copy of the sequence in the target region (b), enabling subsequent allelic analysis. □

Fred Russell Kramer is at the Department of Molecular Genetics, Public Health Research Institute, 455 First Avenue, New York, New York 10016, USA. Paul M. Lizardi is at the Centro de Investigacion sobre Ingenieria Genetica y Biotecnologia, Universidad Nacional Autonoma de Mexico, Apartado Postal 510-3, 62270 Cuernavaca, Morelos, Mexico. For more information, fill in reader service number 100.

1. Harper, M.E., Marselle, L.M., Gallo, R.C. & Wong-Staal, F. *Proc. natn. Acad. Sci. U.S.A.* **83**, 772-776 (1986).

# AIDS conference

At next week's Fifth International AIDS Conference in Montreal, Canada, a host of new products for AIDS research and diagnosis will be on exhibit.

A SYNTHETIC HIV-1 *tat* gene essential for virus growth is the latest addition to British Bio-technology Ltd's line of designer genes (*Reader Service No. 101*). The *tat* gene codes for the transactivator protein of HIV, and its effects are mediated through the transacting response element, which is located within the first 80 nucleotides of the HIV genome, normally transcribed to form the 5' end of viral RNA. Because transgenic mice carrying the *tat* sequence have been shown to develop skin lesions resembling Kaposi's sarcoma, the synthetic gene is potentially useful in developing animal models for HIV research as well as for HIV gene expression and control.

An impermeable plastic backing combined with an absorbent paper surface creates an improved bench cover and spillage confinement sheeting, announces Beta Medical & Scientific (*Reader Service No. 102*). Beta-sorb bench cover can be used for radio-chemical confinement, absorbing routine laboratory chemical spills and as a protective bench covering. Beta says the product's high absorption capacity reduces contamination risks from dripping and eliminates the need for messy tissues. Beta-sorb comes in 46 × 57 cm sheets, which cost £14.97 (UK) for a 50-sheet package and in 46 × 50 m rolls for £28.98 (UK).

## Opportunistic infections

Dako's new monoclonal antibody against *Pneumocystis carinii* stains homogeneous rings corresponding to individual cyst walls (*Reader Service No. 103*). The antibody can be used to stain the antigen in frozen or formalin-fixed, paraffin-embedded lung tissue, in cytocentrifuge preparations of bronchoalveolar lavage fluid, and in sputum samples. Because the antibody does not react with normal tissue, Dako says it presents fewer problems with interpretation. For 1 ml of the antibody, the cost is \$128 (US).

Gene-Trak has announced the availability of its new test for the detection and measurement, in research, of human cytomegalovirus (CMV), (*Reader Service No. 104*). The Gene-Trak CMV assay measures CMV-specific nucleic acids — both RNA and DNA — and provides an estimate of the amount of virus present and the level of virus activity. The kit (\$995 (US) for 100 tests) is a DNA probe-based hybridization assay intended for research applications.

## In situ tests

Using RNA probes and *in situ* hybridization, Oncor says its new HIV-1 *in situ* detection kit is capable of detecting one infected cell out of 400,000 uninfected cells (*Reader Service No. 105*). Specifically, the assay detects the presence of HIV-1 in whole white blood cells as soon as the virus begins to replicate. The <sup>35</sup>S-labelled probes enter the white blood cells and combine with the AIDS virus; the procedure does not require DNA extraction and results are obtained in just over one day. The Oncor kit comes with the reagents and probes to process, hybridize and counter-stain 15 control and 35 sample slides. The kit is priced at \$795 (US).

The new Gene-Trak HIV-1 RNA assay is a DNA probe-based hybridization assay intended for the detection of HIV-1 specific RNA in infected cells in research applications (*Reader Service No. 106*). The hybridization strategy in the assay involves two types of nucleic acid probes: an <sup>125</sup>I RNA detector probe to generate signal and a series of capture probes to retrieve hybrid complexes from solution onto a solid phase. Purification of specific hybridization complexes uses an approach called target cycling/background reduction. In target cycling, hybrids are captured from solution between an unhybridized nucleic acid sequence on the capture probe and a complementary nucleic acid sequence attached to mag-

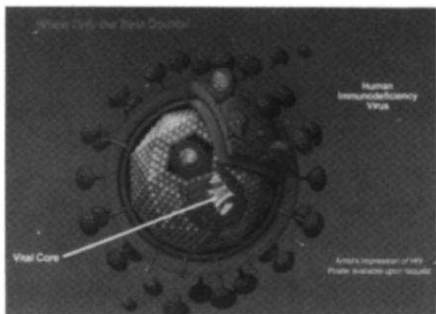
- Leary, J.J., Brigati, D.J. & Ward, D.C. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4045-4049 (1983).
- Chu, B.C.F., Kramer, F.R. & Orgel, L.E. *Nucleic Acids Res.* **14**, 5591-5603 (1986).
- Lizardi, P.M., Guerra, C.E., Lomeli, H., Tussie-Luna, I. & Kramer, F.R. *Bio/Technology* **6**, 1197-1202 (1988).
- Haruna, I. & Spiegelman, S. *Science* **150**, 884-886 (1965).
- Haruna, I. & Spiegelman, S. *Proc. natn. Acad. Sci. U.S.A.* **54**, 1189-1193 (1965).
- Levisohn, R. & Spiegelman, S. *Proc. natn. Acad. Sci. U.S.A.* **60**, 866-872 (1968).
- Miele, E.A., Mills, D.R. & Kramer, F.R. *J. mol. Biol.* **171**, 281-295 (1983).
- Mills, D.R., Axelrod, V.D., Tussie-Luna, I., Lizardi, P.M. & Kramer, F.R. (in preparation).
- Muesing, M.A. *et al. Nature* **313**, 450-458 (1985).
- Zuker, M. & Stiegler, P. *Nucleic Acids Res.* **9**, 133-148 (1981).
- Nishihara, T., Mills, D.R. & Kramer, F.R. *J. Biochem.* **93**, 669-674 (1983).
- Lomeli, H., Tyagi, S., Pritchard, C.G., Lizardi, P.M. & Kramer, F.R. *Clin. Chem.* (in the press).
- Thompson, J. & Gillespie, D. *Anal. Biochem.* **163**, 281-291 (1987).
- Morrissey, D.V. *et al.* (in preparation).
- Shore, D., Langowski, J. & Baldwin, R.L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4833-4837 (1981).
- Rosenberg, M. & Kramer, F.R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 984-988 (1977).
- Okayama, H. & Berg, P. *Molec. cell. Biol.* **2**, 161-170 (1982).
- Milligan, J.F., Groebe, D.R., Witherell, G.W. & Uhlenbeck, O.C. *Nucleic Acids Res.* **15**, 8783-8798 (1987).

netic particles. Gene-Trak says the resulting hybrid can be dissociated under conditions that do not permit dissociation of either capture probe or label probe from the HIV-RNA target. A package of 100 assay tests sells for \$995 (US).

### Target: HIV antigens

Wellcome Diagnostics' Innatest **HIV antigen assay** reacts with p24 and other viral proteins from HIV-1 and HIV-2 and offers a 2.5-hour protocol with only two incubations (*Reader Service No. 107*). The test can be used to detect antigen in human serum, plasma and cell culture supernatant. Based on a microwell format, the Innatest assay is a sandwich EIA that utilizes nonisotopic and non-carcinogenic reagents that have a shelf-life of up to 8 months when stored at 2-8 °C.

Coulter has developed an **HIV antigen immunoassay** to detect p24 and other core-related proteins (*Reader Service No. 108*). The company says the assay has a sensitivity of greater than 10 pg ml<sup>-1</sup> and is specific to all HIV strains. Colour-coded reagents are provided to minimize pipet-



Coulter's view of HIV.

ting error and allow internal quality control. A lyse buffer is used to eliminate infectivity of HIV in samples during the first 15 seconds of the procedure.

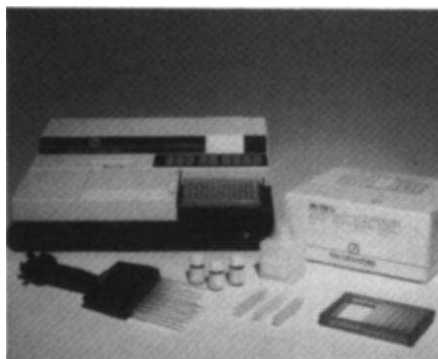
Beckman Instruments is announcing its IgPure line of **five monoclonal antibodies** for the detection of HIV-1 proteins in Western blot and ELISA analysis (*Reader Service No. 109*). Specifically, the antibodies react with viral proteins p24, p17, reverse transcriptase, gp41 and gp120. Beckman says they do not cross-react with other processed HIV proteins. The IgPure products are purified ascites packaged in 100-µg quantities and lyophilised to maintain product stability for no less than 12 months.

### Target: HIV antibodies

Organon Teknika developed its Vironostika anti-HTLV-III assay for the **detection of HIV antibodies** in serum and plasma (*Reader Service No. 110*). The test procedure prescribes a 1 in 34 dilution of the test samples with sample diluent included in the kit's complete set of reagents. The serum or plasma is incubated

in antigen-coated polystyrene Micro-ELISA strip plates. Subsequently, a conjugate of goat anti-human IgG is added, which has been labelled with horseradish peroxidase. The labelled antibody becomes bound by any solid-phase antigen-antibody complex previously formed, and incubation with enzyme substrate produces an orange-yellow colour in the test well.

Flow Laboratories has launched a new version of its HIV-TEK G kit for the detection of **antibodies against HIV** (*Reader Service No. 111*). The new IgG antibody-detecting kit has an increased specificity and sensitivity, says the company. The kit contains 24 eight-well strips coated with purified, inactivated HTLV-III, propagated in T-lymphocyte cell line



HIV-TEK G kit from Flow Labs.

H9. Anti-HIV antibodies present in the tested serum bind to the solid phase and are detected by adding anti-IgG human serum conjugated to peroxidase. Enzyme activity is measured by adding a colourless chromogenic substrate.

### HIV-2 detection

Innogenetics is announcing their new multiple parameter **strip HIV test** based on line immunoassay (LIA) technology (*Reader Service No. 112*). The INNO-LIA HIV-1/HIV-2 antibody test uses a single test strip to detect antibodies to multiple HIV-1 and HIV-2 antigens. Antigens are applied as discrete lines, and no electrophoresis is necessary. Results of the two-hour test can be read by the naked eye.

DuPont has an ELISA kit for the **detection of HIV-2 antibodies** in human serum or plasma (*Reader Service No. 113*). DuPont says the assay is highly specific for HIV-2 and may be used in conjunction with its HIV-2 Western blot kit. DuPont's HIV-2 one-plate ELISA kit sells for \$650 (US). □

*These notes are compiled by Cary Prince from information provided by the manufacturers. To obtain additional information about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.*

## Custom DNA

Purified and Delivered in 48 hours

100 µg, \$10.00 a base for the first 15 bases, \$7.50 for each additional base. Milligram quantities available.

### Research Genetics

2130 S. Memorial Parkway Huntsville, AL 35801

1-800-533-4363

In the United Kingdom 0-800-89-1393

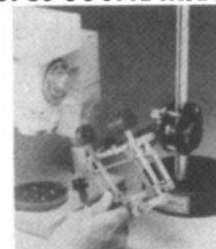
In Canada 1-800-833-8348

FAX 205-536-9016 (INW2971)D

Reader Service No.15

## SINGER INSTRUMENTS

XENOPUS OOCYTE INJECTION?



### SINGER Mk1 MICROMANIPULATOR PLUS INJECT + MATIC

Details from Singer Instruments Co Ltd:  
Roadwater, Watchet, Somerset, England.  
Tel: (0984) 40226. Telex: 337492 Comcab G.

Reader Service No.50

CUSTOM-MADE & PURIFIED

## DNA PEPTIDES

\$8/base Quotes available  
100-150 µg 50-300 mg  
48 hour delivery 1-3 wks delivery

GUARANTEED QUALITY & QUANTITY

V.G. CORP. 1764 Oxford St. E.  
London, Ontario  
Canada N5V 3R6

CALL OR FAX: 1-519-453-0005

Reader Service No.10

## Economical Modular Tissue Culture Incubator



• Safe for AIDS research • Reliable, for In-Vitro fertilization  
• Versatile enough for thousands of lab uses

Easy to use - Just flush with gas mixture, seal and place at approx. temp. Perform multiple experiments simultaneously without cross contamination. Thousands in use worldwide.

For information contact: Flow Laboratories, Ltd. or,  
Billups-Rothenberg, Inc. • P.O. Box 977  
Del Mar, CA 92014-0977 USA • (619) 755-3309

Reader Service No.37



# Academic prospects in the 1990s

Richard Pearson

Traditional academic opportunities and freedoms are not the flavour of the nineties in Britain. What, then are the prospects?

EXPENDITURE on research in the United Kingdom appears to be on the rise again although the extent that funding increases in the public sector represent a real resource increase is debatable.

The level of new job opportunities in research is of course not just a function of expenditure. It also depends on how these funds are allocated between capital and staff costs, and the balance going to different types of employment and, in the case of the research councils, the amount going to postgraduate training. To assess future recruitment opportunities account also has to be taken of the age structure of the existing workforce, likely retirements, and movements to other types of work.

Higher education is still the key source of jobs for academics and at present student numbers is the key variable affecting its budget. Although these are forecast to grow in the next few years costs are continuing to be squeezed and staff: student ratios reduced. With further restructuring needed, the universities have been forecasting a further decline in overall staffing on permanent contracts over the period to 1991. As the expectations are for no massive growth in student numbers over the next decade (*Nature* 339, 80; 1989) the likelihood is that staffing levels will at best remain constant during the 1990s with the possibility of further decline.

## Recruitment

Data and models are available which allow us to project future recruitment levels to the universities which in the United Kingdom employ just over 20,000 permanent staff in science-related areas (see figure). To do this we have to make assumptions about overall staffing levels, age specific retirement and wastage rates and combine these with the known age structure, which is decidedly middle aged. Assuming future loss rates follow those of recent years, and no growth in overall staffing levels, we find that recruitment to the universities in the science related areas will grow over the next two decades by a quarter. However, this only represents an increase in job opportunities of about 400 jobs per annum. Indeed in the early 1990s we can only expect intakes to rise by less than 10 per cent to just over 800 per annum with a further increase to nearly 1,000 per annum in the second half of the 1990s.

At a more specific subject level we see recruitment to engineering and tech-

nology rising from about 150 per annum to just over 200 per annum by the end of the century while recruitment to the sciences is likely to grow at a rather faster rate because of an older age structure and likely lower voluntary wastage rates (see figure). The key influence on future intakes is likely to be the nature of retirement policies. Injections of new posts via a new blood type scheme will have to be massive and regular to have a major affect on job opportunities. Unfortunately,

Annual intakes to university science-related posts

|                      | 1986-90 | 91-95 | 96-00 | 01-05 |
|----------------------|---------|-------|-------|-------|
| Total*               | 740     | 825   | 995   | 1184  |
| Engineers/technology | 150     | 175   | 205   | 230   |
| Sciences             | 270     | 310   | 405   | 490   |

\*Totals include medicine; agriculture, administration, business and social sciences.

similar data are not available for the polytechnic sector, but their age structure suggests that they are unlikely to see significant overall increase in recruitment in the short-term.

While the universities have been reducing the number of their permanent staff there has, however, been a corresponding rise in the numbers on short-term contracts who now account for 1 in 3 academic staff. This reflects a number of changes. First, the rising levels of private sector funding, principally from industry and the charities, the latter being particularly important funders of medical related research. Second, the more general switch away from long-term to contract specific funding by the research councils and other public sector bodies. And finally, policy decisions by universities, and polytechnics, to move away from tenured to short-term contracts so that staffing can be adjusted in the light of changing funding priorities.

The research councils between them directly employ over 7,000 research staff in their establishments, the largest employers being the AFRC although they, along with NERC, have seen significant staff reductions in recent years. These are expected to continue into the future and staffing in this sector is expected to have fallen by about a further 10 per cent over the five years to 1991 despite the rising level of their cash grants. Here again, there is a shift to more short-term contract staff to allow for the rising proportion of short-term funding and to enable staffing to be adjusted in the light of changing

funding levels and priorities.

In the government establishments the pattern is one of retrenchment with a general shift toward a more explicit customer contractor relationship and away from core towards shorter-term funding horizons. Much of the limited recruitment that does take place is also likely to be on a contract basis. They are also being set up to compete for private sector contracts and in some cases privatisation.

In the private sector the picture is less clear with more of an overlap between research and development, but it is clear that many companies are funding pure research with, for example, major expansions being announced by ICI and Glaxo in the pharmaceutical sector while Hitachi have just announced a new laboratory linked in with the Cavendish at Cambridge. The flourishing of the science parks is further testimony to the growing private sector involvement in research although the parks embrace a wide range of activities including marketing, production and development as well as research. We are also seeing a growing internationalisation of private sector research with US and Japanese companies setting up in the UK to access Europe, while some of the international electronics firms are setting up in the South of France offering a package of the latest facilities and good location which they believe will be attractive to the best researchers from throughout the world.

## Public and private

European Commission money through programmes such as ESPRIT as well as commercial considerations are also extending links across frontiers and between higher education and the private sector, further blurring organisational boundaries. In contrast to the public sector there is little evidence of a growing use of short-term contracts as these are seen as a deterrent to attracting the best staff.

In conclusion, we see a decline in academic research opportunities in the public sector with many of the new jobs, where they are available, being offered on a short-term contract basis. By way of contrast prospects are improving in the private sector where more permanent opportunities are on offer and with an increasingly international flavour although with potentially less academic freedom. □

Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Falmer, Brighton BN1 9RF, UK.